

AFIT/GSM/SM-78S-3

LEVEL

①

AD A0 65971

DDC  
RECEIVED  
MAR 15 1979  
RECEIVED  
C

DDC FILE COPY

THE PERFORMANCE-POTENTIAL  
DILEMMA--DOES IT EXIST?

THESIS

AFIT/GSM/SM-78S-3 Drew W. Browning  
Capt USAF

Approved for public release; distribution unlimited.

79 03 13 003

14

AFIT/GSM/SM-78S-3

6

THE PERFORMANCE-POTENTIAL  
DILEMMA, DOES IT EXIST?

9

Master's  
THESIS

Presented to the Faculty of the School of Engineering  
of the Air Force Institute of Technology  
Air Training Command  
in Partial Fulfillment of the  
Requirements for the Degree of  
Master of Science

12

1334

10

by

Drew W. Browning, ~~B.S.~~

Capt

USAF

Graduate Systems Management

11

September 1978

Approved for public release; distribution unlimited.

0421 1.1.3

## Preface

The primary reason for this research effort was that it was necessary to meet the requirements for a Master of Science Degree in Systems Management. Additionally, my interest in conducting research on the Officer Evaluation System (OES) was heightened by the fact that the OES has a direct effect on my, as well as every other officer's, future in the Air Force.

This research effort has hopefully provided some additional insights on one aspect of the complex and controversial subject of employee appraisal: the evaluation of individual potential.

Several individuals have made this research effort quite a bit easier and deserve my most sincere expression of appreciation. First of all, I wish to thank Major E. J. Dunne, who eased my anxieties about thesis work greatly by providing me with this research topic and serving as my advisor. Thanks are also in order for the computer help provided by Major Charles W. McNichols and Captain Harold E. Klick. Finally, a very special thank you is expressed to Molly and Bill Bustard, whose courteous and helpful attitudes in the AFIT library and computer center, respectively, made this research effort much more bearable.

Drew W. Browning

ACCESSION for

NTIS ☒ W. A. Section ☒  
 DOC ☐ B. Section ☐  
 UNANNOUNCED ☐  
 JUSTIFICATION

BY  
 DISTRICT

A

## Contents

	<u>Page</u>
Preface . . . . .	ii
List of Figures . . . . .	v
List of Tables. . . . .	vi
Abstract. . . . .	viii
I. Introduction . . . . .	1
Air Force Officer Evaluation . . . . .	2
Previous Research . . . . .	6
Statement of the Problem. . . . .	7
Approach to the Problem . . . . .	8
Objectives . . . . .	9
Scope and Limitations. . . . .	10
Assumptions. . . . .	12
Summary . . . . .	12
II. Research Methodology . . . . .	13
Literature Search. . . . .	13
Policy Capturing: Modeling Human Judgment . . . . .	15
Previous Research in Judgment Modeling . . . . .	17
Statistical Basis for a Policy Capturing Model. . . . .	18
Development of the Decision-making Exercise. . . . .	22
Selection of the Sample Population . . . . .	28
Data Analysis . . . . .	30
Specific Analyses Performed on the Data . . . . .	30
III. Appraisal Systems: An Examination of the Literature . . . . .	36
Historical Development and Trends . . . . .	37
Evolution of the Air Force Officer Evaluation System. . . . .	41
The Objectives of Employee Appraisal. . . . .	45
Formal and Informal Appraisal. . . . .	48
What Do Appraisal Systems Measure? . . . . .	50
Indicators of Performance and Potential. . . . .	55

Contents (continued)

	<u>Page</u>
IV. Results . . . . .	59
Exercise Response . . . . .	59
Demographic Classification of Respondents . . . . .	59
Individual Regression Results . . . . .	62
Paired Sample t-Test of Individual Relative Weights . . . . .	66
Group Regression Results . . . . .	67
F-Test Comparison of Group Regression Models . . . . .	71
Examination of an Interactive Model . . . . .	76
Part III Exercise Analysis . . . . .	77
V. Summary and Conclusions . . . . .	80
Conclusions . . . . .	83
Implications of the Study . . . . .	86
A Final Comment . . . . .	87
Bibliography . . . . .	89
Appendix A: Decision-Making Exercise . . . . .	93
Appendix B: Air Force Form 707 . . . . .	116
Appendix C: Air Force Form 77 . . . . .	119
Appendix D: F-Test Computational Form . . . . .	122
Appendix E: Demographic Frequencies . . . . .	125
Vita . . . . .	141

## List of Figures

<u>Figure</u>		<u>Page</u>
1	Example Scenario . . . . .	26
2	Example Part III Scenario . . . . .	29
3	Performance Model . . . . .	51

# List of Tables

<u>Table</u>	<u>Page</u>
2.1 Scenario Cue Levels . . . . .	27
4.1 Exercise Response Rates . . . . .	59
4.2 Exercise Respondents by Grade . . . . .	60
4.3 Exercise Respondents by Aeronautical Rating. . . . .	60
4.4 Exercise Respondents by History as an OER Rater. . .	61
4.5 Exercise Respondents by Most Recent OER Rating . . .	61
4.6 Individual Regression Results . . . . .	64
4.7 Percentages of Individual Regressions Significant . . .	65
4.8 t-Test of Individual Relative Weights . . . . .	66
4.9 Group Regression Results and a Comparison with Individual Regression Results . . . . .	68
4.10 F-Test Comparison of Group Regression Models (PME School Groupings) . . . . .	71
4.11 F-Test Comparison of Group Regression Models (Aeronautical Rating). . . . .	72
4.12 F-Test Comparison of Group Regression Models (History as an OER Rater). . . . .	72
4.13 F-Test Comparison of Group Regression Models (Recent OER Rating) . . . . .	73
4.14 Group Regression Coefficients (History as an OER Rater). . . . .	74
4.15 Group Regression Coefficients (Recent OER Rating) . .	75
4.16 Comparison of Interactive and Non-interactive Models . . . . .	77

List of Tables (continued)

<u>Table</u>		<u>Page</u>
4.17	Part III Responses (Mean Values and Standard Deviation) . . . . .	78
4.18	Paired Sample t-Test (Part III Data) . . . . .	79



Abstract

Many organizations, including the United States Air Force, utilize employee appraisal systems which require the evaluation of individual potential in addition to the appraisal of one's past performance. With regard to its officer evaluation system, the Air Force has established only broad guidelines that indicate various criteria to be used in the evaluation of potential, one of which is past performance. Thus, each individual officer, either as a rater or as a ratee, makes judgments of OER ratings based on his or her perceptions of what criteria should be used to evaluate potential. It is hypothesized that if widely varied perceptions do exist, many organizational dysfunctions may occur.

This research effort examines Air Force officers' perceptions, both individually and in groups, of what is involved in a rating of potential. In order to determine these perceptions, a judgment modeling, or "policy capturing" research approach was used. A decision-making exercise was administered to 381 active duty USAF officers attending Squadron Officers' School, Air Command and Staff College, and Air War College. In this exercise, officers were asked to judge the appropriateness of OER ratings assigned to 18 hypothetical officers, based on different levels of two factors: (1) the officer's past performance, as illustrated by three indicators and (2) the

officer's potential, as illustrated by three indicators other than past performance. Data collected from the exercise were then analyzed to test various hypotheses concerning the relative influence of performance and (other) potential factors on officers' judgments of OER ratings.

The results of the research indicate that Air Force officers do not view the OER rating as being solely determined by past performance, although the performance factor was weighted most heavily. Analysis indicated that officers used different judgment policies when considering "good" and "bad" OER ratings (potential was weighted more heavily when judgments were made of "bad" OER ratings), and that they were generally internally consistent in their judgment policies. Additionally, several different analyses indicated that the Air War College students used judgment policies that were significantly different than the other groups; specifically, AWC students generally placed greater emphasis on potential than the other PME groups. Furthermore, the AWC group displayed both the lowest internal judgment consistency and the lowest group  $R^2$  values (an

indication of the degree of decision-making homogeneity within the group).

# THE PERFORMANCE-POTENTIAL DILEMMA:

## DOES IT EXIST?

### I Introduction

This research effort is concerned with employee appraisal systems--frequently called performance appraisal, merit rating, personnel evaluation, and a host of other terms that are all related to the evaluation of individuals regarding their work. Although subject matter in this report is discussed in terms of the U.S. Air Force Officer Evaluation System (OES), many of the inferences drawn from this research could have significant implications for similar civilian as well as military personnel appraisal systems.

George Terry, a noted authority on management, has stated that as supervisors, we are constantly judging and forming opinions about the contributions and abilities of our group members. The question, therefore, is not whether to appraise employees, but how to appraise them (Terry, 1974:224). If an organization does not clearly address and answer this question of how to appraise employees, then a number of problems may arise and render the appraisal system ineffective, or only marginally effective, in accomplishing its stated goals.

The subject of employee appraisal has a reputation of generally being looked upon with disdain by managers, even though appraisal

systems are, by design, an aid to management rather than a hindrance. There has therefore been a great deal written about the problems associated with appraisal systems, and there are a number of "solutions" to these problems presented in the form of recommended types of, or variations to, appraisal systems.

One source of problem associated with appraisal systems relates to the question of what the system is designed to evaluate, and subsequently how well this "design" corresponds to what is actually evaluated in practice. Appraisal systems that are designed to evaluate and/or predict more than simple job performance (for example, an estimation of an individual's future worth as well as his or her past performance) may result in differing perceptions among raters and ratees regarding what is actually being evaluated and what degree of importance each factor has in determining an individual's overall appraisal rating. One such appraisal system that is designed to evaluate more than job performance is the Air Force Officer Evaluation System.

#### Air Force Officer Evaluation

A widely discussed and controversial topic among Air Force officers today is the officer evaluation system. Even though changes in the evaluation system in recent years (e.g. forced distribution ratings and the recent revision of distribution quota percentages) have heightened the controversy and interest, the basic fact remains that

any system which affects officers as much as the officer evaluation system will always generate a great deal of personal interest and concern. In order to more closely address the question of why officers are so intensely concerned with the OES, one need only examine the intended objectives of the system. According to Air Force Regulation 36-10, Officer Evaluations, the objectives of the officer evaluation system are as follows:

The officer evaluation system must provide the Air Force with essential information for use in personnel decisions, such as promotions, assignments, and school selections. The system permits better identification of officer quality differences, and assists in identifying and motivating officers for due course and accelerated promotions (AFR 36-10, 1978:1-1).

Thus, the officer evaluation is not intended to be a simple performance appraisal apparatus which seeks only to judge an officer's past performance relative to other officers being rated: an officer's future worth is also "evaluated," again relative to other officers being rated. The OES leads to key decisions regarding an officer's career progression and is therefore a primary factor in determining his or her future in the Air Force.

Two related but distinct concepts which form the basis of the Air Force officer evaluation system are those of performance and potential. The Air Force Form 707, "Officer Effectiveness Report," (OER), has two evaluation sections: Part III involves a rating of performance factors, and Part IV involves an evaluation of potential. This topic will be discussed in more detail in chapter three of this

report, along with a historical development of the current OER system. The current OER form for company grade officers is located in Appendix B.

Since the officer effectiveness report (as an instrument of the officer evaluation system) requires an assessment of both performance and potential, one might argue that the individuals involved with the OES should be able to clearly distinguish between these concepts if the appraisal system is to effectively accomplish its desired objectives as outlined in pertinent Air Force regulations. It should be pointed out that this distinction affects every Air Force officer, either as a rater or as a ratee.

Performance can be defined as the degree to which an employee has satisfactorily accomplished his or her job or assigned duties during a specified period of time, and is determined primarily by one's abilities, coupled with motivational forces. Measuring performance may be thought of by many as a simple process of identifying an individual's units of work output. However, many jobs, including a large proportion of high level managerial and professional jobs, simply have no readily identifiable units of output; in these instances, the appraiser observes the employee's performance and records the observations with some form of appraisal instrument (Cummings, 1974:4).

Potential, on the other hand, is a concept which is related to the future. Although potential is not always predicted in appraisal

systems, it is more prevalent in appraisal systems for management personnel. For the purpose of the officer evaluation system in the Air Force, potential is defined in AFR 36-10 as follows:

Potential is the ratee's capability, relative to that of officers in the same grade in the group being evaluated, for expanded/more diverse responsibility. Potential for the purpose of OER rating decisions will be determined primarily by the evaluator's assessment of the ratee's accomplishments during the period of the report with consideration given to other whole person factors, e.g. career pattern, integrity, PME, etc. (AFR 36-10, 1978: 2-3).

Since potential is future oriented, measurement of this individual capability can be a difficult task. Generally speaking, raters will examine past performance, trends in performance, and personality traits in order to estimate future performance. AFP 36-26 indicates that job performance is normally the single most accurate indicator of potential (AFP 36-26, 1977:3-1). It is noted, however, that the degree to which other factors determine this rating of potential is not addressed in Air Force policy.

The concepts of performance and potential are discussed definitionally in Air Force regulations, but not specifically addressed are the inherent problems involved with a failure by raters and/or ratees to clearly distinguish between these concepts when actually making judgments of various OER ratings (Section IV rating, "Evaluation of Potential"). More specifically, since potential appears to be difficult to assess in terms of various traits, abilities, and motivation, there may be a natural inclination for raters to place a heavy (if not total)

emphasis of performance factors when judging an individual's potential.

### Previous Research

There have not been many studies conducted that have examined the distinction between the concepts of performance and potential with respect to judgments of appraisal ratings. The relationship between these concepts and the necessity of a clear distinction between the two was, however, specifically addressed in one study of the officer evaluation system:

"The basis of many past problems with military officer evaluation systems is that seldom, if ever, have past systems clearly distinguished between performance and potential, i. e., have not separated appraisal of the immediate past performance from appraisal of future performance. Performance appraisal and potential appraisal are different processes which use different input information and whose results should have different uses." (Dunne, 1977:22, 32)

Additionally, one thesis research effort which examined junior officer perceptions of the OER system resulted in the author's conclusion that junior officers feel as though job performance is the most important factor making up their potential. The author additionally pointed out that whether or not job performance can be the prime measure of potential is open for debate (Keyserling, 1976:63). This research effort further investigates this subject area by attempting to answer the question of whether officers distinguish between the concepts of performance and potential when making judgments of OER ratings, as well as whether or not officers are consistent in making



these judgments.

#### Statement of the Problem

Due to the interrelated nature of performance and potential, as well as the difficulties involved in assessing the two factors in an officer evaluation system, there is an ever-present possibility that an individual's concepts of performance and potential could become confused and overlap significantly when actually making judgments of OER ratings. It should be noted that this confusion is not to be interpreted to mean that officers are not able to define or recognize definitions of performance and potential. Rather, it is confusion regarding the meaning and implications of the results of the officer evaluation system. For example, the Air Force officer evaluation system appraisal instrument provides for a rating which is an "evaluation of potential" (Section IV, Air Force Form 707). It is hypothesized that this rating has strong performance implications, ones that may have not necessarily been intended. This research examined the strength of these performance implications as an indicator of whether confusion does exist regarding the assessment of potential in Air Force officers. Another indication of confusion that this study examined was the variability among officers' judgment policies regarding OER ratings.

If this confusion exists in the minds of raters or ratees, the effectiveness of the officer evaluation system in distinguishing between

the two will be limited to individual perceptions. These varied individual perceptions may subsequently have dysfunctional effects that could undermine the intended goals of the officer evaluation system. A discussion of the effects of such perceptual differences is found in a 1977 article titled "An Integrated Approach to Control System Design" (Ansari, 1977). Ansari proposed that successful control requires leaders who can create conditions which motivate workers toward desired goals; and subordinates must share, or at least must not reject, their manager's perceptions of what the goals are, how they are to be met, and how well they were met (evaluation). Ansari makes the following statement about evaluation:

"The greatest and potentially most harmful (perceptual) differences are likely to be at the evaluation stage. In fact, different perceptions of means and ends may well affect a subordinate's performance and thus show up in evaluation also." (Ansari, 1977:109)

#### Approach to the Problem

This research effort investigates Air Force officers' perceptions of performance and potential in terms of the influence that each concept has on judgments of OER ratings. In order to evaluate these perceptions, a policy capturing approach was used, along with an additional survey measure of officers' more intuitive reactions to either receiving or assigning both "good" and "bad" OER ratings. The nature and use of policy capturing techniques in investigating rater policies will be discussed in a subsequent chapter of this report.

## Objectives

The overall objective of this research was to investigate the relationships between officers' concepts of performance and potential and their judgments of appraisal (OER) ratings.

Specific objectives that were pursued in support of the overall thesis objective were as follows:

1. Develop a theoretical background for the research subject by conducting a literature search on employee appraisal systems, with particular emphasis on the role of performance and potential assessment in fulfilling the goals of appraisal systems.
2. Develop and administer a decision-making exercise that "captures" the policies of Air Force officers regarding their judgments of OER ratings, based on performance and potential criteria.
3. Analyze the data collected through the decision-making exercise in order to determine the relative weights that officers place on performance versus (other) potential criteria when making judgments of OER ratings.
4. Analyze the data collected through the decision-making exercise in order to determine the consistency with which officers make judgments of OER ratings based on performance and potential criteria, both individually and as groups.
5. Examine the correlation between officers' intuitive reactions to OER ratings (Part III of the exercise) and the results of the policy capturing analysis (Part II).

The data obtained in the research was used to investigate the following research hypotheses:

H1: Air Force officers combine performance and potential criteria in an additive rather than interactive manner when making judgments of OER ratings. (Note: This hypothesis facilitates the policy capturing data analysis.)

H2: Air Force officers, when considering (1) job factors which are indicators of immediate past performance and (2) personal factors which are (other) indicators of potential, will make judgments of OER ratings (Section IV, "Evaluation of Potential") based equally on the job and personal factors.

H3: Air Force officers use the same judgment policies when judging either "good" or "bad" OER ratings (i. e. "1" and "3," respectively) in terms of the performance and potential factors that influence these judgments.

H4: Air Force officers are internally consistent in their judgment policies concerning OER scores.

H5: Air Force officers represent a homogeneous group of decision makers regarding OER scores; i. e., use the same decision policies when making judgments of OER ratings, regardless of data groupings by PM<sup>7</sup> school, aeronautical rating, OER rating history, and whether or not officers have been OER raters.

H6: The relative weights for performance and potential criteria determined by the policy capturing model are consistent with the intuitive responses given in Part III of the exercise; i. e., modeled judgment policies are consistent with officers' intuitive reactions.

#### Scope and Limitations

This study involved an initial search of the literature on the subject area of employee appraisal systems and a subsequent investigation of the perceptions of Air Force officers regarding the concepts of performance and potential. It is emphasized that this research effort was not intended to be a critique of the officer evaluation

system. Additionally, it was not an attempt to determine what importance should be associated with various predictors of potential. Rather, investigation was directed at empirically determining whether officers distinguish between the concepts of performance and potential when making judgments of OER ratings and the consistency with which officers make these judgments based on performance and potential criteria.

Although this research was aimed at determining the judgment policies of Air Force officers in general, only a small sample could be selected to participate in the decision-making exercise. For this reason the sample population was restricted to students at the three Professional Military Education (PME) schools at Maxwell AFB, Alabama. The exercise was pretested at the Air Force Institute of Technology.

One limitation associated with the decision-making exercise was the selection of "indicators" of performance and potential: although many other factors could have also been included, the desire to retain the simplicity of the exercise resulted in the selection of three representative indicators of performance ("job factors") and three of potential ("personal factors"). The rationale for the selection of these factors is presented in chapter three of this report.

Finally, the researcher acknowledges the difficulties inherent in attempting to quantify officers' decision-making policies regarding OER scores: in actual practice, an OER score is assigned based on

many factors, some of which may not be directly related to the concepts of performance and potential. This research was therefore designed to examine the influence of performance and potential on officers' judgments of OER ratings, with all other factors being equal.

### Assumptions

The assumptions under which this research was conducted and upon which conclusions were based are as follows:

1. Responses to the decision-making exercise reflected honest perceptions of the respondents and not "school answers," or a gaming of the exercise.
2. Job factors and personal factors selected for use in the decision-making exercise represent significant aspects of the concepts of performance and potential, respectively.

### Summary

In summary, this initial chapter has introduced the research subject, identified the research problem, and has outlined the objectives, limitations, and assumptions of the study. In the remaining chapters, the methodology employed in investigating the research questions is presented, followed by a presentation of research results.

## II Research Methodology

The methodology of this study was designed to address the research objectives outlined previously, and was conducted in the following sequence:

1. A literature search on appraisal systems and the concepts of performance and potential was conducted.
2. A decision-making instrument was developed to "capture" judgment policies of Air Force officers concerning performance and potential criteria in appraisal system decisions.
3. A sample population of Air Force officers was selected.
4. The decision-making exercise was administered to the sample population.
5. Collected data was analyzed to evaluate stated research hypotheses.
6. Conclusions and recommendations were made based on analysis results.

### Literature Search

This research was concerned with employee appraisal systems-- an area of managerial activity that every Air Force officer has experience in dealing with, either as a rater (of other officers) or as a ratee. However, each officer may view the officer evaluation system differently in terms of perceived evaluation system goals, the inputs to the system, and the methods used in administering the system.

Since this study examined officers' perceptions of the two appraisal concepts known as performance and potential, a theoretical

background on the subject area of appraisal systems was needed. The development of a theoretical background was therefore included in this research effort to serve as (1) background information for developing the nature of the research problem, as applied to both civilian and military appraisal systems, and (2) a theoretical basis with which officer perceptions of performance and potential could be compared.

Several sources were used in this assimilation of information on appraisal systems. The AFIT library provided a readily accessible source of many periodical references, as well as information on military applications of appraisal techniques. The Wright State University library provided an excellent source of personnel administration books that were devoted to employee appraisal. Additional information was obtained from pertinent Air Force publications and government research studies obtained through the Defense Documentation Center (this particular source was extensively utilized to obtain information on policy capturing techniques).

Valuable information on Air Force research of appraisal systems was obtained from the Air Force Human Resources Laboratory (AFHRL), Brooks AFB, Texas, particularly a 1971 AFHRL report titled USAF Officer Evaluation Systems Review and Research Recommendations (AFHRL:1971).

Apart from military appraisal systems, information on one company's program to identify executive potential was obtained



through personal contact with individuals in the Employee Relations Office of the Exxon Corporation.

It is noted that although the literature search additionally involved a survey of available information on policy capturing (or judgment modeling) and its application in determining appraisal decision policies, this subject was considered to be part of the research methodology and is therefore discussed next in this chapter.

#### Policy Capturing: Modeling Human Judgment

Since a rater's involvement with the officer evaluation system entails making judgments of other officers, a policy capturing, or "judgment modeling" technique may be employed to construct a mathematical model of the process by which these judgments are made. From this model, one may be able to draw inferences about the influence that various factors or "cues" have on the decision process as well as the consistency with which "judges" make these decisions.

In a general sense, policy capturing has been defined by one researcher as

. . . the identification and quantification of the attributes that are pertinent to a decision and the subsequent mathematical description of the design policy for the evaluation of these attributes (Gooch, 1972).

An initial reaction of some people to this concept may be one of skepticism, based on the belief that human judgment processes are too complex to mathematically model. However, an Air Force

research study on policy capturing indicates just the opposite point of view:

" . . . most human decisions are not complex . . . the variables in a decision may be complex, but when left to the limited skills of the human mind, the situation will be simplified to a manageable surrogate. [Policy Capturing] systematically and efficiently externalizes this process." (Jones, et al, 1975).

In relation to employee appraisal systems, the "policy of a rater" has been defined as "what raters do" when they are asked to respond to a series of stimuli, and "capturing" a rater's policy is defining the process of predicting the actions of that rater from the known characteristics of the stimuli he or she is being required to evaluate (Naylor and Wherry, 1965:969).

The need for understanding how raters process and evaluate information was emphasized by one researcher who suggested that statistical techniques be used to determine more clearly the basis of policy decisions, and that the weights actually given to criterion elements in making judgments of performance should be determined and systematized. Such techniques, it was argued, would determine the implicit weighting, for a judge or rater, of the factors contributing to assessment of others and would permit a comparison of an individual's implicit policy with the explicit weighting policy adopted by the organization (Smith, 1976).

The value, therefore, of policy capturing is that it permits the inference of a rater's weighting policy by requesting assessments of

research study on policy capturing indicates just the opposite point of view:

" . . . most human decisions are not complex . . . the variables in a decision may be complex, but when left to the limited skills of the human mind, the situation will be simplified to a manageable surrogate. [Policy Capturing] systematically and efficiently externalizes this process." (Jones, et al, 1975).

In relation to employee appraisal systems, the "policy of a rater" has been defined as "what raters do" when they are asked to respond to a series of stimuli, and "capturing" a rater's policy is defining the process of predicting the actions of that rater from the known characteristics of the stimuli he or she is being required to evaluate (Naylor and Wherry, 1965:969).

The need for understanding how raters process and evaluate information was emphasized by one researcher who suggested that statistical techniques be used to determine more clearly the basis of policy decisions, and that the weights actually given to criterion elements in making judgments of performance should be determined and systematized. Such techniques, it was argued, would determine the implicit weighting, for a judge or rater, of the factors contributing to assessment of others and would permit a comparison of an individual's implicit policy with the explicit weighting policy adopted by the organization (Smith, 1976).

The value, therefore, of policy capturing is that it permits the inference of a rater's weighting policy by requesting assessments of

a total object, or an overall evaluation of a ratee, rather than requiring explicit evaluation of elements. The focus of the analysis is on the decision of the rater rather than on the rater's interpretation of his or her decision processes (Zedeck and Kafry, 1977:287).

#### Previous Research in Judgment Modeling

There has apparently been very little research conducted on how raters actually combine, process, or integrate information in forming overall assessments of their ratees. Three studies in this subject area, however, are known to the researcher.

Naylor and Wherry (1965) used a regression analysis for policy capturing in a performance appraisal situation. Each of 50 Air Force supervisors rated each of 250 profiles (ratees) described in terms of 23 traits, where each trait had a possible 1-9 range. The rater then judged a global criterion judgment on each ratee, and a regression algorithm was used to describe the decision processes.

Taylor and Wilsted (1974) used a linear regression technique for capturing raters' policies, where the raters were Air Force Academy cadet officers. Results indicated that the raters were internally consistent in their policies but varied greatly with each other.

Finally, a policy capturing technique was used in a study where 67 nurses evaluated (as raters) 40 hypothetical nurses who were described by nine criterion elements, each having three possible levels of performance (Zedeck and Kafry, 1977). Overall evaluations

were made on a seven point scale, ranging from "very effective" to "very ineffective." The researchers' analysis yielded relative weights for each of the nine criterion elements; i.e., the influence of each element on the overall evaluation was determined, as well as the consistency with which raters made their decisions.

The usefulness of policy capturing in relation to employee appraisal is not restricted to an evaluation of already existing systems: Kavanagh, MacKinney, and Wolins (1971) reviewed the content issue in performance appraisals; i.e., the researchers proposed that prior to implementing a rating program for performance appraisal, judgments of the content or stimuli to be rated must be made. The content is viewed as potentially ranging from subjective to objective, abstract to concrete, or personality to performance. Thus, Kavanagh proposes the use of a policy capturing technique in the actual planning of an appraisal system.

#### Statistical Basis for a Policy Capturing Model

The statistical approach used in this study is based on Hoffman's proposal that it is possible to model the human judgment process by representing the policy of a decision maker as a linear combination of stimulus variables, or "cues" (Hoffman, 1960). Regression analysis is then used to describe a judge's decision with the information leading to that decision by means of correlational statistics (Jones, et al, 1975).

The basic approach requires the rater to make quantitative evaluations of a number of ratees, each of whom is defined by one or more criterion elements, each of which in turn has several possible levels. If there are a sufficient number of ratees for whom the rater must provide an overall judgment, and each of the ratees is described by the same set of elements (though different levels), then a regression equation can be developed for a rater which describes his or her idiosyncratic method of combining and weighting information. The mathematical model represents the strategy or policy of the rater; the relationship between each criterion element and the overall assessment, the beta weight, identifies the relative degree of importance of the criterion elements as they influence the rater's decisions.

The regression model takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \epsilon$$

Y is the (dependent) variable that represents the decision being modeled; the  $\beta$ s are parameters of the regression model; the  $X_i$ s are the predictor variables (stimuli, or "cues" in policy capturing); and " $\epsilon$ " represents an error term (accounts for any variance not explained by the predictor variables).

By fitting empirical data (various values of the decision variable Y for given levels of the predictor variables ( $X_i$ )) to a linear equation by the least squares regression model, the (standardized)  $\beta_i$ s can be determined.

Hoffman (1960) points out that the standardized beta weights signify the importance attached to each of the predictor variables by the judge. Large coefficients mean, empirically, that the corresponding predictors can account for a large proportion of the variance of judgment; conversely, a predictor with a small beta weight contributes little beyond the contribution of other predictors. Hoffman points out, however, that there are limitations to the "characterization" of the judgment process by means of beta weights: beta coefficients do not account for all of the predictable variance and do not allow for the assessment of the relative contribution of each predictor. He then argues that a set of weights are desired which are theoretically capable of accounting for all of the predictable variance and which carry exact interpretations in terms of components of variance (Hoffman, 1960:119-120).

Therefore, in order to determine the relative influence of each predictor variable, or cue, on the decision made by a judge (i.e., a corresponding value of the decision variable "Y"), Hoffman defined an index termed "relative weight." It is defined as follows:

$$RW_{iy} = \frac{\beta_{iy} r_{iy}}{R_y^2}$$

$RW_{iy}$  is the relative weight of predictor  $X_i$ ;  $\beta_{iy}$  is the standardized regression coefficient (beta weight) of predictor  $X_i$  regressed on the decision variable Y;  $r_{iy}$  is the intercorrelation coefficient between the

predictor variable  $X_i$  and the decision variable  $Y$ ; and  $R_y^2$  is the squared multiple correlation coefficient of the regression equation. Hoffman emphasizes that "relative weight" provides a means of portraying the relative contributions of each of the predictors such that a simple sum of them accounts entirely and unambiguously for the predictable variance (Hoffman, 1962:78).

If the policy capturing technique involves the use of orthogonal predictors in the regression the beta weights and intercorrelation coefficients are identical; thus, the relative weight equation reduces to the following:

$$RW_{iy} = \frac{\beta_{iy}^2}{R_y^2}$$

The orthogonality issue does not imply that the predictors are independent (i.e., no interaction), but rather refers to the design nature of the instrument. Jones, et al, point out that the requirement for design orthogonality forced those involved with the investigation of human judgment modeling to turn to factorially designed experiments in order to obtain orthogonal predictors (Jones, et al, 1975:19). This research involved the use of a policy capturing instrument with orthogonal predictors, the development of which will be discussed later in this chapter.

Thus, the policy capturing analysis yields the relative degree to which each predictor variable influences the rater's decision, and



the calculated  $R_y^2$  values are interpreted as measures of the consistency of judges in applying decision policies (Jones, et al, 1975:18). It is noted that both individual and group decision policies can be mathematically modeled using the policy capturing technique, although the concept of relative weight is meaningful only in the case of individual models. Also, " $R^2$ " as a measure of consistency applies to both individuals and groups, although one would expect group  $R^2$ s to be lower than individual  $R^2$ s, since raters may be internally consistent in their decision policies yet differ widely with other raters (Madden, 1963:2).

#### Development of the Decision-making Exercise

Part I of the exercise was designed to collect standard demographic data on respondents such as grade, aeronautical rating, and OER rating history. The demographic data base was established in order to examine possible correlations between these factors and a judge's decision-making policies.

Part II of the decision-making exercise was designed to describe a linear judgment model of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

The Decision Variable. Y is the decision, or exercise response value, which indicates a respondent's judgment (on a five point Likert scale) of the appropriateness of an OER rating assigned to a

hypothetical officer. This method allows judgments to be "captured" relative to overall perceptions of factors that may determine an OER rating, regardless of whether the judge views the situation from a rater or ratee point of view. An alternative approach was to have respondents assign OER scores to hypothetical officers; it was determined, however, that the total rating process in the Air Force is quite complex--involving several raters and perhaps requiring several iterations before a final rating is decided upon. It would have therefore been very difficult to develop a brief exercise which simulated this process.

A simpler "decision" was selected to investigate the relative weights of the two predictors (performance and potential) in judgments of OER scores. The decision was as follows: given a hypothetical officer with an identified level of "performance," as well as the level of some other indicators of "potential," judge the appropriateness of a given OER rating (the rating in the OER section "Evaluation of Potential"). Respondents were then asked to select one of five responses on a five-point Likert scale.

Predictor Variables.  $X_1$  and  $X_2$  are predictor variables which represent the following:

$X_1$  - Performance of a ratee, as described by three indicators ("job factors") of job performance during the reporting period.

$X_2$  - Potential of a ratee, as described by three indicators ("personal factors") of potential other than job performance.

This exercise utilized three indicators of each predictor since one can not easily describe a level of performance or potential with one simple statement. Rationale used in selecting these indicators is discussed in chapter three.

Three levels were selected for each predictor variable ( $X_1$  and  $X_2$ ); i. e., all three "indicators" for each predictor were either (1) good, (2) neither good nor bad, or (3) bad.

The specific indicators used for each predictor (performance and potential) were as follows:

$X_1$ : Job Factors (indicators of performance)

1. Accomplishment of assignments
2. Efficiency in the use of resources
3. On-the-job communicative ability and effectiveness

$X_2$ : Personal Factors (indicators of potential, exclusive of job performance)

1. Motivation and dedication
2. Intellectual ability
3. Sensitivity to people and their needs.

Again, it is emphasized that three levels were used for each predictor "set," and these levels were assigned values of +1 ("good"), 0 ("neither good nor bad"), and -1 ("bad").

The exercise was designed to represent a factorially designed experiment; i. e., all possible combinations of predictor variables (performance and potential) are represented. Thus, design

orthogonality was ensured and the concept of relative weight is applicable.

Exercise Format. Cases were constructed by combining each level of the  $X_1$  variable with all possible levels of the  $X_2$  variable. Since there were three levels for each of the two cues, there were  $3 \times 3$ , or 9 combinations possible. However, since the OER rating assigned to each scenario was given as either being a "1" ("good") or a "3" (considered "bad" under the most current OER system), there were  $3 \times 3 \times 2$ , or 18, cases which were presented. The nine cases corresponding to each OER score were analyzed separately to determine if the policies of judges varied when considering "good" or "bad" OER ratings. An example of the decision-making exercise scenario used in this research is presented in Figure 1.

An additional assumption made in this exercise to prevent confusion in the minds of the respondents was that there was agreement among the rater, additional rater, and reviewer regarding the OER rating given for each scenario. This assumption simplified the decision to be made, in terms of removing the possibility of respondents considering different levels of OER review.

The 18 scenarios were ordered randomly in the exercise by using a random number table (Freund, 1971:445), and correspond to the following levels for each predictor variable and the OER rating assigned (Table 2.1).

OFFICER #1

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION #1

How appropriate is the OER rating that was given? (Circle one number.)

1	2	3	4	5
Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate

Figure 1. Example Scenario

Part III of the exercise was not part of the policy capturing technique, but rather was an attempt to gauge respondents' reactions to four commonly held beliefs regarding what an OER rating means (two were performance-related and two were potential-related). The

Table 2.1  
Scenario Cue Levels

Scenario Number	OER Rating	Job Factors (Performance)	Personal Factors (Potential)
1	1	Low	High
2	3	Low	Low
3	1	Medium	Low
4	3	Low	High
5	1	High	Low
6	3	High	Medium
7	3	Medium	Medium
8	3	High	High
9	3	Low	Medium
10	3	Medium	High
11	1	Low	Low
12	1	Medium	Medium
13	1	Low	Medium
14	1	Medium	High
15	3	Medium	Low
16	3	High	Low
17	1	High	Low
18	1	High	Medium

purpose of Part III was to obtain additional data relating OER score, performance, and potential in order to make comparisons with the policy capturing data. The Part III data is more direct and conceptually simple than the policy capturing information, and may be considered more of an intuitive reaction of officers to OER ratings.

These four "beliefs" about OER scores were presented in the context of four different situations: receiving a "1" rating, receiving a "3" rating, assigning (as a rater) a "1" rating, and assigning a "3" rating. The four areas were related to (1) promotability, (2) effectiveness at higher level jobs, (3) how well the job was done, and (4) job

performance relative to others. The first two beliefs relate to potential, while the second two concern job performance.

Respondents were then asked to give their reactions to each of the four statements on a five point scale indicating agreement or disagreement. An example of one of the four OER situations used in this part of the exercise is presented in Figure 2. As in Part II, statements were ordered randomly in the actual exercise.

The complete decision-making exercise used in this research is located in Appendix A.

#### Selection of the Sample Population

The necessity for having a sample population that was readily accessible and convenient to survey, resulted in the following sample population composition of active duty Air Force officers:

1. A random sample of 186 Squadron Officer School (SOS) students (Lieutenants and Captains).
2. A random sample of 204 Air Command and Staff College (ACSC) students (Captains and Majors).
3. All (USAF) Air War College (AWC) students (194) (Lieutenant Colonels and Colonels).

One might argue that the sample population has an inherent bias in that the schools selected are generally regarded as having students who are "above average" officers. The researcher acknowledges this possibility, yet it is pointed out that if inconsistencies in judgment policies are found for "above average" officers, then there is intuitively an even greater chance of average, or below average,

A. You have just received an OER rating of "3." For each of the possible reactions listed below, circle the number which best describes your feeling toward each reaction.

1. They (my boss, my chain of command, the "system") don't think I should be promoted.

1	2	3	4	5
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

2. They are saying that most other people performed their job better than I did mine.

1	2	3	4	5
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

3. They don't think I could be effective in higher level jobs.

1	2	3	4	5
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

4. They are saying I didn't do my job very well.

1	2	3	4	5
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

Figure 2. Example Part III Scenario

officers exhibiting judgment inconsistencies.



### Data Analysis

Analysis of collected data was conducted with the use of the Aeronautical Systems Division CDC 6600 computer, Wright-Patterson AFB, Ohio. Descriptive statistics programs, group regressions, and paired sample t-tests incorporated in the analysis were accomplished using the Statistical Package for the Social Sciences, known as SPSS (Nie, et al., 1975).

Individual regression equations and associated statistics were formulated using a Fortran program written by Captain Harold E. Klick, also a graduate systems management system. This program calculated, in addition to individual regression equation coefficients, the relative weights for each predictor variable and the multiple correlation coefficient ( $R^2$ ) for each exercise respondent. This information was then stored on a disk file for use in later computations.

### Specific Analyses Performed on the Data

frequencies on Exercise Responses. In order to provide a method for examining the raw data base, the SPSS program "frequencies" was used (Nie, et al., 1975:194) to provide frequency distributions on the responses to each exercise question. This output provided a concise descriptive analysis of the exercise respondents according to grade, aeronautical rating, and other demographic variables, as well as the distributions of the decision exercise responses.

Individual Regression Analyses. Two regression analyses were performed (using the Fortran program) with the data from each exercise respondent: one regression was accomplished using the nine scenarios wherein the assigned OER was a "1," and the second regression was accomplished using the remaining nine scenarios in which the assigned OER score was a "3." In each regression, the degree of appropriateness of the assigned OER (the exercise response) was regressed on the two decision cues: (1) performance and (2) other indicators of potential. For each OER "set," this program yielded the relative weights that each officer placed on performance and potential factors, as well as the  $R^2$  value associated with the decisions made. Mean values of these parameters for different groups of officers were calculated and are presented as research results.

Paired Sample t-Test of Individual Relative Weights. In this research, there was not an a priori assumption that officers use a single decision-making policy when making judgments of both "good" (a "1" rating) and "bad" (a "3" rating) OER ratings. Rather, the two sets of scenarios were treated separately (as described in the preceding section), and a paired sample t-test (Nie, et al, 1975:272) was conducted on the two sets of data in order to determine if mean values of relative weights were different for the two given OER levels. This analysis therefore indicated whether officers used different decision-making policies in judging "good" and "bad" OER ratings.

Group Regression Analyses. In addition to individual decision-making policies, this research also involved an examination of the judgment policies of groups of officers. In this regard, the degree of appropriateness of the given OER score (the decision variable) was regressed on the performance and potential factors, based on the total number of decisions made by the following groups of officers:

1. All groups (381 respondents)
2. SOS students (113)
3. ACSC students (144)
4. AWC students (124)

As in the case of the individual regression analyses, separate regressions were run on the scenarios with the "1" and "3" ratings. The different combinations of groups were regressed in order to provide information to test the hypotheses that various groups used different judgment policies about OER scores. This analysis, which was accomplished using an F-test for comparison of regression models, is discussed in the next section of this chapter.

The output of the group regression analyses yielded standardized beta weights for each predictor (performance and potential), the group  $R^2$ , and analysis of variance (ANOVA) data used in later F-test computations.

F-Test for Comparison of Regression Models. The F-test, as described by G. C. Chow (1960:599) was used to test whether there was a statistically significant difference (at the .05 level) in

regression models being compared. The null hypothesis in each comparison states that there is no significant difference in the regression coefficients of the models being compared, versus an alternate hypothesis that there is at least one model among those being compared whose regression coefficients are different. When more than two models are being compared, however, this test does not identify which model(s) are different; therefore, when a null hypothesis is rejected in such a case, the number of models being compared in subsequent analysis may be partitioned into smaller groups in order to more closely identify which models differ from others. This approach was used in this research effort, as indicated in the results chapter of this report.

The F-test of significance was used not only to compare regression models based on the PME school of respondents, but also involved hypothesis testing with groupings of respondents based on demographics such as aeronautical rating, recent OER rating, and respondents' history as an OER rater.

#### Examination of Interactive versus a Non-interactive Model.

Although this research was primarily involved with judgment modeling based on an additive effect of predictor variables on the decision variable, some researchers have proposed that higher order, or "interaction," terms should also be included in the models (Jones, et al, 1975:21). In essence, these higher order terms (e.g., an  $(X_1 \times X_2)$  term) may be added to the regression model in order to account for

interaction between predictor variables, therefore adding explanative power to the model. Since there were two predictors used in this research, one interaction term (performance level times the potential level) was added to each regression model in order to determine the statistical significance of the interaction term, as well as the resultant increase in group  $R^2$ . By examining the relative contribution of interaction terms, one may then make observations about the "completeness" (or lack thereof) of the non-interactive model in predicting judgments.

Paired Sample t-Test on Part III Data. Part III of the exercise was included to obtain more direct reactions to OER scores with which the results of the policy capturing model could be compared.

As described earlier in this chapter, there were four situations presented in Part III: awarding a "1," receiving a "1," awarding a "3," and receiving a "3." Following each situation, there were four commonly held beliefs stated about OER scores with which the respondent was asked to indicate either agreement or disagreement (on a five point scale). In each situation, two of the "beliefs" were performance oriented, and two were potential oriented.

In order to determine which concept received the most "agreement," a paired-sample t-test was conducted to examine whether the agreement with either potential beliefs or performance beliefs was greater than the other. It was hypothesized that the relative weights

calculated from the policy capturing model should be consistent with this additional analysis of the Part III data.

### III Appraisal Systems: An Examination of the Literature

As discussed previously, this research was concerned with the Air Force officer evaluation system, and the possibility of officers viewing the system differently in terms of perceived evaluation system goals, as well as inputs to the system (performance and/or potential evaluation).

Although this study dealt with a military appraisal system, it involved many concepts that also apply to civilian appraisal systems. The following discussion of appraisal system development and concepts is presented in order to serve as background information for clarifying the nature of the research problem, and to provide a theoretical basis with which officer perceptions of performance and potential (as determined by analysis efforts) can be compared.

The first topic discussed is the historical development of appraisal systems in general, as well as the apparent trends in appraisal techniques that have been documented in the literature. A similar discussion of the Air Force officer evaluation system is then presented. Following this historical development of civilian and Air Force appraisal systems, the objectives of employee appraisal are described, followed by a discussion of what appraisal systems measure. Finally, a representative list of "indicators" of both performance and potential is developed, as supported by management literature. These

indicators were used in the decision-making exercise that formed the basis of this research effort.

### Historical Development and Trends

The managerial activity known as employee appraisal is not a recent product of the "technological age." References to appraisal of managers can be found in antiquity: emperors of the Wei dynasty (AD 221-265) had an "imperial rater" whose task it was to evaluate the performance of the official family. Centuries later, Ignatius Loyola established a system for formal rating of the members of the Jesuit Society (Whisler & Harper, 1962:423).

One of the first uses of formal appraisal systems in the United States was by government agencies in the late 19th century, a move induced by criticism of waste and the spoils system in government. The real impetus to appraisal in business, however, came as the result of the work of Frederick Taylor and his followers before World War I. These appraisal systems were related to various efficiency factors developed from work simplification and time and motion study (Koontz, 1971:17).

As concern with human relations aspects of the work environment increased in the 1930s and 1940s, behavioral traits such as "ability to get along with others" tended to become increasingly prevalent in appraisal systems.



Early appraisal systems were generally aimed at the evaluation of hourly workers rather than managers. Koontz points out that managerial appraisal systems largely introduced during World War II and immediately thereafter had their foundation in hourly labor performance appraisal systems. Too many of them, he argues, were based on worker qualities and attitudes rather than on performance (Koontz, 1971:19).

One source pointed out that management literature, especially since World War II, has been increasingly concerned with the problem of performance appraisal; in a 1962 collection of 50 reprint articles on performance appraisal, only 8 were written before 1951. "This suggests," the authors pointed out, "that although performance appraisal is not new, it has only recently gained the attention it deserves" (Whisler & Harper, 1962:2).

With rising interest in behavioral science following World War II, there was a tendency toward a traditional "trait approach" to employee appraisal. One study, conducted in 1957, concluded the following:

"Most companies . . . are concerned mainly with personality and character traits . . . so strongly is the emphasis on personality that job knowledge and even job performance may have only a minor place in the overall rating" (Dale and Smith, 1957:22).

There have been many explanations offered for this "drift" toward such a trait approach: one analyst offered three underlying reasons: (1) as behavioral sciences captured public interest, there was a greater tendency to explain effectiveness by psychological or

psychiatric measures (2) since so many managerial tasks cannot be given quantitative measurement and since qualitative factors are difficult to measure, there has been a tendency to drift toward personality-centered appraisals (3) the third factor relates to the manager's job; practically every study has found successful managers to be strong leaders and have highlighted human relations skills (Kellogg, 1965:60).

Beginning in the late 1950s and early 1960s, the merits of employee appraisal were widely argued. Douglas McGregor, in his widely read article "An Uneasy Look at Performance Appraisal," condemned the traditional approach to appraisal wherein managers were put in the "untenable position of judging the personal worth of subordinates, and then acting on these judgments." (McGregor, 1957:94) Other criticisms of traditional appraisal systems were offered by Likert (1959), Stolz (1960), Mayfield (1960), and others.

McGregor's alternative approach to appraisal, which places the major responsibility on the subordinate for establishing performance goals and self appraising progress toward them, appears to have its basis in Drucker's concept of management by objectives, or "MBO." Under such an approach, an employee and his or her supervisor periodically negotiate what the subordinate should accomplish by the end of the rating period (Drucker, 1954:121).

The inclusion of MBO concepts in appraisal systems is becoming increasingly prevalent: a recent study indicated that over half of 293 firms surveyed used some form of MBO in their appraisal system.

The study concluded that MBO-based appraisal methods, since its measurements are clearly job-related and appraisals are easier to make: " . . . an individual either meets his objectives, or he doesn't." (Anderson, 1978:1).

The move to objective or results-oriented appraisal has also been precipitated by equal opportunity laws: recent federal court rulings essentially have made it the responsibility of an employer to prove that his rating system does not discriminate; companies must now demonstrate both that their systems are reliable and that the factors that are measured are truly related to job performance. Actual court cases have upheld employee complaints of "patently subjective" judgments based on factors that had little to do with how the job was done. Anderson thus concludes that as lower-level executives become increasingly anonymous in the expanding white-collar work force, many companies do not want to rely on informal rating systems: they believe they will be in a better position to avoid conflict with equal opportunity laws if they justify personnel decisions with uniform appraisal standards (Anderson, 1978:1).

Concomitant with the increasing reliance on formal appraisal, however, is increasing criticism of most systems. There appears to be instability regarding many organizations' appraisal systems: a recent report by the Conference Board noted that over half of the 293 firms it surveyed had developed new systems within the past three years, yet current systems were still widely regarded as a nuisance

at best and a dangerous evil at worst. One consultant admitted that finding a workable appraisal system "seems like a search for the Holy Grail." (Anderson, 1978:1)

#### Evolution of the Air Force Officer Evaluation System

The history of the Air Force officer evaluation system is rather short, but nevertheless illustrates the evolutionary process that many managerial appraisal systems undergo.

When it became a separate service in 1947, the Air Force continued to use the Army officer evaluation system, which had been substantially revised earlier in 1947. Whereas the Army system had previously been based on a trait approach to appraisal, the "new" AGO Form 67 required a different method of evaluation: the essence of this form was one section which provided a rater with multiple-choice questions about ratee characteristics, and an additional section which provided space for written comments by the rater (Rhoades, et al, 1978: 10).

General dislike, however, of this system prompted the Air Force to implement its own OER system with the introduction of Air Force Form 77, 15 March 1949. This system, although based on extensive research of officer perceptions of rating factors by the American Institute of Research, was also found to be lacking in that it was too complex (there were 54 items to be scored), time consuming, and gave rise to "inflated" ratings (AFR 36-10, 1975:A-2).

In 1952, a new system was introduced that "survived," with minor modifications, for 22 years. Two somewhat similar forms were used with this system: the Air Force Form 77 (used for rating company grade officers) was task oriented, while the Air Force Form 707 (used for rating field grade officers) was management oriented (Rhoades, et al, 1976:12).

The front side of both forms was used for identification data (Section I), and a description of duties (Section II). Section III described eight "rating factors," which were listed in terms of a five block scale ranging from "outstanding" to "poor" descriptors. Section IV entailed a similar five block rating of "military qualities." Section V, which was titled "Overall Evaluation," involved a ten block scale ranging from "unsatisfactory" to "outstanding."

Section VI involved a rating of "Promotion Potential," on a four block "descriptor" scale. Finally, Section VII provided space for rater comments, and an additional section provided space for comments by an indorsing official. The Air Force Form 77, November 1966, can be found in Appendix C.

Although this system lasted for 22 years, problems with rating "inflation" led to its demise. In 1961, less than five percent of all Air Force Officers were receiving "top block" ratings in the "Promotion Potential" section: in 1974, over 90 percent had received top ratings five consecutive times (Rhoades, et al, 1976:1).

In November 1970, the Director of Personnel at Headquarters USAF requested that the Air Force Human Resources Laboratory (AFHRL) develop a new officer evaluation system, based on the belief that the current system was inadequate for differentiating between individuals for selection (promotion) and assignment purposes (Johnson, et al, 1976:5).

The proposed OER system was designed by the AFHRL around a management by objectives concept: the system was based on a statement of job objectives and was designed to assure some degree of counseling in the rating process. Additionally, a separate evaluation of promotion potential was recommended, since AFHRL researchers questioned the capability of a "single system doing everything for everybody." (Johnson, et al, 1976:5, 8). The AFHRL recommendations specifically stated that the Air Force CES objectives should be to produce valid and discriminating assessments of (1) performance and (2) potential, in order to provide information for assignment, counseling, and selection (AFHRL, 1971:App6). Additional emphasis on potential as well as performance assessment was stressed in the "personnel plan," approved by the Secretary of the Air Force, which stated that promotion is to be based on performance and potential--in the past, emphasis has been almost entirely on past performance (AFHRL, 1971:Tab2, App 2).

The OER review group, which considered the recommendations of the AFHRL, substantially modified the proposed system: the MEO

and the "two form" concepts were eliminated, as was the requirement for interaction between the rater and ratee regarding the appraisal of job accomplishments (Johnson, et al, 1976:10,11).

More specifically, the current OER system does not outline this "developmental" function as an objective of the system: AFR 36-10 states the following:

"Evaluation reports are designed for the personnel management of Air Force Officers. They are not to be used as counseling devices . . . counseling is performed (continuously) by the supervisor for the purpose of assisting individuals in improving their performance." (AFR 36-10, 1978:1-2)

The revised OER system was implemented by the Air Force in November 1974. The new Air Force Form 707, November 1974 has two evaluation sections: Part III involves a rating of performance factors similar to that found in the previous system except that it has no overall evaluation of performance, and Part IV is titled "Evaluation of Potential." This evaluation of potential is scored in terms of a six block scale ranging from "highest" to "lowest." In order to control rating inflation in this section, a controlled distribution was imposed on the top two blocks of the rating of potential (AFR 36-10, 1975:5-3). For clarification purposes, a "top block" rating in Section IV is commonly referred to as a "1," the second level block a "2," etc. (through "6," the lowest rating possible).

In January, 1978, the controlled rating distribution was modified to reflect limitations only on the percentage of top block ratings

awarded (AFR 36-10, 1978:6-1). Since the current controlled rating distribution only applies to the "evaluation of potential," some writers have keyed on the assessment of potential as the essence or heart of the current OER, such as Blakelock (1976:30), Rhoades, et al (1972:5). and Dunne (1977:12). The concept of "potential" assessment versus "performance" assessment leads to an examination of some of the problems associated with appraisal systems. The remainder of this discussion will therefore focus on the sources of these problems: the objectives of appraisal systems, and the question of what the systems are designed to appraise.

#### The Objectives of Employee Appraisal

Peter Drucker, in his discussion of the need for employee appraisal, pointed out that "insistence on high goals and high performance requires that a man's ability both to set goals and to attain them be systematically appraised." He additionally pointed out that since managers make many decisions based on appraisals of employees, these managers need a systematic appraisal or else too much time is wasted on making decisions by "hunch" rather than by knowledge (Drucker, 1954:149).

Kellogg states that to appraise anything is to set a value on it; however, there is no such thing as a universally accepted value. She additionally proposes that managers should discard the notion that appraisals are "absolute" evaluations with which associates will agree:



he or she should recognize, instead, that appraisal is a subjective judgment made on the basis of information which is incomplete (Kellogg, 1965:15). This inherent subjectivity is reflected in the following definition of employee appraisal:

"Employee appraisal is a manager's subjective judgment of the value of an individual's ability to do something: the manager judges one's present and/or future usefulness to a job or a business." (Kellogg, 1965:19)

Additionally, Kellogg points out that employee appraisal is most likely to be sound if its purpose is well defined (Kellogg, 1965:19). A discussion of the objectives of employee appraisal, therefore, is presented at this point.

In any managerial activity, one must initially consider the goals of that activity. In this regard, an examination of the goals of employee appraisal is in order. This discussion is considered essential, as one source indicated that a fundamental reason for the failure of performance appraisal systems is the failure of management to clearly define objectives and to establish techniques specifically designed to accomplish these objectives (Huse, 1967:3).

In general, the primary purpose of employee appraisal is to facilitate improved results. In an attempt to accomplish organizational goals, managers are actively involved in direct contact with subordinate managers and workers by executing the "directing" and "controlling" managerial functions. In this regard, employee appraisal results in an "information base" on employees that is used for two distinct purposes.

The first of these is personnel administration, wherein the manager evaluates subordinates regarding their performance (and possibly potential for future performance), and subsequently forms the basis for decisions concerning promotions, transfers, and merit rewards, including salary administration (Cummings, 1973:4).

In order to "improve results" in the short run, managers evaluate the performance of subordinates and take necessary actions to promote better performance. In the long run, however, managers generally consider (and evaluate, at least mentally) both the potential of an employee (future worth to the organization) and the development of employees to fulfill future needs of the organization. The concepts of performance and potential will be examined in more detail in a subsequent section of this chapter, so discussion will now focus on the second use of appraisal information: personal development.

From this viewpoint, appraisals are designed to help improve performance or the capability for performance directly by aiding the employee in identifying areas for improvement and growth (Cummings, 1973:4-5). This developmental function is essential to both the subordinate (for personal growth and increased job competence) and the organization. If, through the appraisal process, managers are successful in further developing an employee's effectiveness, organizational goal accomplishment will have been facilitated.

This developmental approach therefore emphasizes the importance of managerial actions in improving and facilitating an individual's

effectiveness relative to his or her own abilities (Cummings, 1973:6). The manager is therefore placed in a counseling role in his or her relationship with subordinates. Terry emphasizes that it is the manager's job to guide subordinates, not to inspect and drive them toward sought for improvements (Terry, 1974:226).

Employee appraisal therefore establishes an information base upon which personnel administration and personal development actions are hopefully taken to facilitate organizational goal accomplishment.

#### Formal and Informal Appraisal

Although many managers may think of employee appraisal in terms of formal systems (i. e., documented appraisals on a specified form at required intervals), there also exists a continuous, informal type of appraisal. The manager, through day-to-day contacts with subordinates, continuously appraises their performance and takes actions which hopefully improve performance. Informal appraisal can be thought of as one input to the formal appraisal of subordinates. Thus, the informal appraisal of subordinates involves maintaining effective communications lines between superiors and subordinates that serve as routes for positive feedback as well as for corrective measures.

The two perspectives for viewing appraisal--personnel administration and personal development--are applicable to both formal and informal means of appraisal. Whereas the personnel administration viewpoint is most often associated with formal appraisal systems, the

personal development viewpoint is most frequently associated with the manager's day-to-day, informal contact with subordinates. In some appraisal systems, however, personal development is additionally accomplished (at least by design) through a formal appraisal interview: these appraisal interviews are conducted periodically and involve a discussion between the supervisor and the subordinate of the formal performance rating, strengths and weaknesses, and areas needing improvement. Thus, the appraisal interview serves as an important source of feedback to the employee that lets him or her know where he or she stands.

Although there may be much to be gained from an informal system of employee appraisal, it appears that many managers are either reluctant or ineffective in this regard. Thus, major emphasis is placed by many organizations on the formal system of appraisal, particularly with respect to personnel administration and associated decisions regarding salary, promotions, etc. This research was oriented toward the implications of formal appraisal systems, particularly as related to the Air Force officer evaluation system.

A significant question to be asked when considering appraisal systems concerns what the system is intended to measure. This subject is certainly essential to any effective appraisal system, for managers must be aware of and understand what is being evaluated before the appraisal system can be effective in achieving desired results.

### What Do Appraisal Systems Measure?

This topic concerns what appraisal systems are designed to measure, as well as some of the obstacles inherent in attempt to subjectively evaluate human performance (or the capability for performance). Particularly when dealing with formal appraisal systems, managers must be cognizant of what is being measured; if managers actually evaluate employees based on factors different than those intended by the appraisal system design, dysfunctional organizational effects are possible which limit the accomplishment of organizational goals.

Performance. In many appraisal systems, supervisors are simply evaluating employees' past performance (i.e., performance appraisal). Since the concept of performance is being discussed, a reiteration of the definition presented in chapter one is in order:

Performance is defined as the degree to which an employee has satisfactorily accomplished his or her job or assigned duties during a specified period of time.

Performance is determined primarily by one's abilities, coupled with motivational forces; i.e., one must have both the capability and the desire to do it. However, Cummings points out that since performance is an individual phenomenon, environmental variables influence performance, primarily through their effect on the individual determinants of performance--ability and motivation, as shown in the model below.

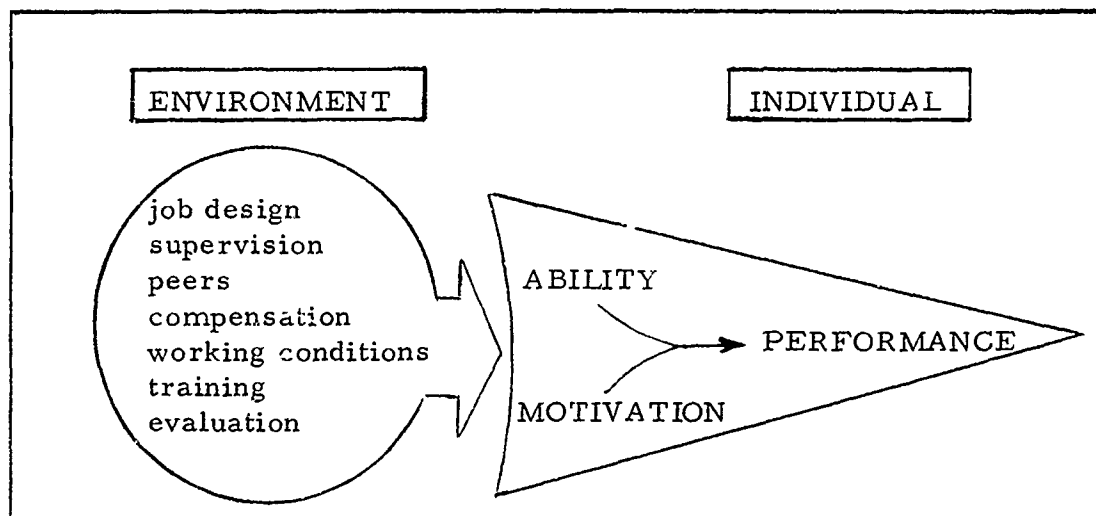


Figure 3. Performance Model (Cummings, 1973:9)

Ability reflects capability; a relatively stable characteristic which enables persons to behave in some specified fashion. Motivation, on the other hand, reflects effort or energy, a dynamic characteristic which determines how vigorously capabilities will be employed in some activity (Cummings, 1973:8).

One might argue that examining the effects of environmental factors on the determinants of performance is not relevant to the subject of appraisal, since performance is viewed in retrospect (past actions are considered). However, if some of these environmental factors have adverse effects on the determinants of performance, then one's demonstrated level of performance may not be truly indicative of an employee's true capabilities to perform, i. e., that employee's potential. Patz points out that past performance is only one indicator of what future performance may be at higher levels. Additionally, he

emphasizes that promotion decisions, like all decisions that affect the future, should reflect estimates of management potential more strongly than appraisals of past performance (Patz, 1975:79). Laurent emphasizes that performance in one job is, to a degree, predictive of performance in another, but generally only to the extent to which the two jobs are similar; the accuracy of the prediction decreases as the jobs become more dissimilar (Laurent, 1968:3).

Potential. As was alluded to in earlier discussion, managers are not only concerned with past performance, but are additionally oriented toward the effective performance of subordinates in the future, particularly when these subordinates are also managers. Some appraisal systems are therefore designed to predict the future worth--the potential--of individuals to the organization.

One private industry research report emphasized a primary reason for the . . . st in identifying executive, or management, potential:

"The job of a manager in a business organization is becoming more and more complex, which means that those going into management positions in the future must have more ability than some of those who were able to do an acceptable job in the past. These needs put a premium on the early identification of employees who have the potential to be successful in management positions."  
(Laurent, 1968:1)

One survey of top and middle managers in 19 companies indicated that the purpose of appraisal is not only a developmental technique to improve results, but is also a necessary vehicle for

assessing management potential (Patz, 1975:75).

Not all management writers support the appraisal of individual potential, as evidenced by Drucker's statement that an appraisal which "focuses on potential, on personality, on promise, on anything that is not provable performance, is an abuse." (Drucker, 1954:150) Although in 1954 openly opposing the rating of potential in employee appraisal systems, Drucker appears to lend some support (possibly unintentional) to the appraisal of potential in his more recent book,

The Effective Executive:

"Effective executives know that they have to start with what a man can do rather than what a job requires. This means that they do their thinking about people long before the decision on filling a job has to be made, and independently of it. The purpose is to arrive at an appraisal of a man before one has to decide whether he is the right person to fill a bigger position." (Drucker, 1967:83)

Drucker's comment appears to indicate that one cannot always wait on proven performance to base certain decisions; one therefore makes predictions of one's future performance based on various factors.

As was mentioned in earlier discussion, one of the objectives of employee appraisal is to provide information upon which promotion decisions are based. One study indicated that some companies placed major weight in promotion decisions on the immediate superior's estimate of the subordinate's potential, rather than on the superior's estimate of the subordinate's job performance effectiveness (Campbell, et al, 1970:37).



Since potential is future oriented, a manager may encounter difficulties when attempting to measure this individual capability. These difficulties are illustrated by the following excerpt from an Exxon USA research report on measuring management potential:

"If the evaluation of performance is difficult, the evaluation of potential is even more so. This is particularly true when one attempts to evaluate the management potential of young employees." (Laurent, 1968:2)

Kellogg indicates that the first and most important thing for a manager to do is recognize his or her limitations in predicting the future and do everything possible to gather accurate, relevant, representative information as a basis for judgments. Therefore, the manager makes a judgmental decision about the probability that a given employee will grow at a rate which will place him or her in competition for certain key openings in the future. Kellogg also points out that because of the relative inadequacy of this decision, it should be considered a short term one, subject to frequent reviews as the future unfolds both for the firm and for the individual (Kellogg, 1965:137). Drucker also supports this "short term" concept by indicating that one cannot appraise potential for any length of time ahead or for anything very different from what a man is already doing (Drucker, 1967:85).

Regarding Air Force officer evaluations, the emphasis also appears to be on a "short term" evaluation of potential. AFR 36-10, in discussing potential, emphasizes this short range time frame:

"Evaluators will focus primarily on near term capabilities for increased responsibilities; from an evaluator's perspective, long range assessment in a subordinate element for which primary responsibility rests with the selection board." (AFR 36-10, 1978:2-22)

Thus, employee appraisal systems may be designed to measure (past oriented) performance and/or (future oriented) potential in support of appraisal systems objectives. The next section of this chapter presents several "indicators" of performance and potential, as supported by various sources in the management literature.

#### Indicators of Performance and Potential

As was mentioned in earlier discussion, measuring the performance of managers is not always a simple task, particularly since a large proportion of managerial and professional jobs simply do not have readily identifiable units of output. Thus, it appears that the higher the job level in the organizational hierarchy, the more abstract and difficult to measure job performance becomes. Furthermore, the evaluation of management potential appears to pose even greater difficulties since potential is based on predictions of future performance, rather than exclusively on assessments of past performance.

Since this research involved a policy capturing method of determining officer perceptions of performance and potential, "indicators" of each concept were chosen for inclusion in the decision-making exercise. In general, performance represents one's past accomplishments in terms of effectiveness and efficiency, while potential is a more abstract concept that is additionally based on various personality traits

and individual capabilities. The following three indicators of performance are considered a representative, but certainly not comprehensive, listing based on current management literature.

### Performance Indicators

1. Accomplishment of assignments--This indicator, selected to represent the degree of on-the-job accomplishments during a rating period, is perhaps the most obvious performance indicator and is supported, among others, by Pigors and Myers (1977:47, 274).

2. Efficiency in the use of resources--Not only is performance determined by how effective a job is accomplished, but also by the degree of efficiency with which resources are utilized. This reasoning is supported by Mann and Dent (1954:105), Johnson, Meehan and Wilkinson (1976:7), and efficiency is included as a performance factor on the Air Force Officer Effectiveness Report (Air Force Form 707).

3. Communicative ability and effectiveness--The importance of being a proficient communicator, both orally and in written form, has been emphasized by Kellogg (1965:142), Johnson, et al (1976:7), and is also included as a performance factor on the OER form.

Three representative indicators of potential obtained from a search of management literature (indicators other than immediate past performance) are as follows:

### Potential Indicators

1. Individual motivation and dedication--This "drive" factor is listed in AFP 36-26, paragraph 3-2c, e as "essential and without which an officer's potential is strictly limited." (AFP 36-26, 1977) Additional support for this factor is offered by Guyton (1969:659), and Laurent (1968:3).

2. Intellectual ability--Guyton concludes that this factor is a strong contributor to success prediction (Guyton, 1969:660). Other sources which support this factor are Cummings (1973:16) and Laurent (1968:3).

3. Sensitivity to people and their needs--AFP 36-26, paragraph 3-2J specifically points out that human relations is an important aspect of potential. Koontz (1971:22) lends additional support to the inclusion of human relations ability as an indicator of management potential.

Although recent performance and trends in past performance are considered to be valid predictors of potential, these factors were not included in the decision-making exercise as indicators of potential in order to clearly distinguish between "performance-related" and "other" indicators of potential. This rationale was considered essential to the research effort, since it has been argued that in some places the frame of reference in military appraisal systems is "potential," but the assumption is that potential is little more than performance (Dunne, 1977:12). In order to examine this argument, a policy capturing model was developed to distinguish between the relative importance that

officers place on (1) past performance and (2) other indicators of potential when making judgments of OER ratings (Section IV, "Evaluation of Potential"). The resulting weights that officers placed on performance and potential were used to examine if potential is, in fact, assumed in some situations to be "little more than performance," as Dunne hypothesizes.

#### IV Results

##### Exercise Response

As indicated in Table 4.1, a total of 584 decision-making exercises were distributed to students at the three PMF schools at Maxwell AFB, Alabama. Of the exercises returned, 381 were usable, resulting in an overall return rate of 65.2%.

Table 4.1  
Exercise Response Rates

PME School	Number of Exercises Distributed	Number of Usable Returns	Return Rate
SOS	186	113	60.8%
ACSC	204	144	70.6%
AWC	194	124	63.9%
Total	584	381	65.2%

##### Demographic Classification of Respondents

The 381 respondents represented a wide range of grade, aeronautical rating, command of assignment, years of service, education level, and other demographic variables. Tables 4.2 through 4.5 illustrate respondent classification by grade, aeronautical rating, history as an OER rater, and most recent OER rating. Additionally, 16 different commands were represented in the survey group, the

Table 4.2  
Exercise Respondents by Grade

Grade	Absolute Frequency	Relative Frequency (%)
0-2	23	6.1
0-3	107	28.1
0-4	127	33.3
0-5	95	24.9
0-6	29	7.6
Total	381	100.0

Table 4.3  
Exercise Respondents by Aeronautical Rating

Aero Rating	Absolute Frequency	Relative Frequency (%)
Pilot	164	43.0
Navigator	49	12.9
Non-rated	168	44.1
Total	381	100.0

mean education level was a master's degree; and 87% of the respondents indicated that they either definitely or most likely will make the Air Force a career.

Table 4.4  
Exercise Respondents by History as an OER Rater

History as a Rater	Absolute Frequency	Relative Frequency (%)
Never rated another officer	117	30.7
Occasionally rated another officer	112	29.4
Frequently rated another officer	150	39.4
Missing Cases	2	.5
Total	381	100.0

Table 4.5  
Exercise Respondents by Most Recent OER Rating

OER Rating	Absolute Frequency	Relative Frequency (%)
"1"	192	50.4
"2"	112	29.4
"3" through "6"*	71	18.6
Other or Never had an OER	5	1.3
Missing Cases	1	.3
Total	381	100.0
*Only one officer had an OER rating of "4" or less.		



An interesting result of one demographic question was that 31% of the exercise respondents had never written an OER on another officer; this factor was later used in investigating statistical differences in regression models. Also, the distribution of the responses to question 14 ("What rating did you receive on your most recent controlled OER?") indicated that a majority of officers had received "1s" (specifically, 50.5% versus a 22% Air Force average). Furthermore, only 18.7% responded that they had received a "3" or less, as opposed to an Air Force average of 50% under the 1974-1977 OER system. Assuming valid responses to these questions, the students at the three PME schools received higher OER ratings than Air Force officers in general.

A complete listing of responses to the 15 demographic questions is found in Appendix E.

#### Individual Regression Results

As outlined in the methodology chapter of this report, two regression analyses were performed on the data for each exercise respondent in order to construct individual models where the "degree of appropriateness" of an OER rating was the decision (dependent) variable, and the factors "performance" and "potential" were the predictor (independent) variables.

Using the previously mentioned Fortran program, two regression models were constructed for each respondent: one for the nine scenarios with an assigned OER rating of "1," and one for the nine

scenarios with an OER rating of "3." The regression algorithm calculated the standardized regression coefficients, relative weight for each predictor, and squared multiple correlation coefficient ( $R^2$ ) for each individual's pair of regression models.

In order to more readily present the results of these individual regressions, mean relative weights and mean  $t$  values were calculated for the groupings of officers by PME school and are presented in Table 4.6. Table 4.7 lists the percentages, by PME group, of the individual regressions that were significant at the .05 level.

A preliminary inspection of the relative weights assigned gives the appearance of general agreement among groups of officers regarding the performance and potential criteria used in judging OER ratings. However, specific hypothesis tests were conducted later in the analysis using group regressions to determine if the group regression models were statistically different.

It is noted that "performance" as a decision criterion consistently had a higher relative weight than "other indicators of potential" for all of the PME groups. An interesting result of the separate regressions for "1" and "3" scenarios was that for all groups, the relative weight for potential was higher for the "3s" scenarios model than for the "1s" scenario model. This implies that potential was weighted more heavily when these officers made judgments of "bad" OERs than it was when they made judgments of "good" OERs.

Table 4.6  
Individual Regression Results

Respondent Group	Scenario OER	Predictor	Mean Relative Weight	Mean $R^2$
SOS	OER="1"	Performance	.70	.686
		Potential	.30	
	OER="3"	Performance	.64	.704
		Potential	.36	
ACSC	OER="1"	Performance	.73	.737
		Potential	.27	
	OER="3"	Performance	.62	.713
		Potential		
AWC	OER="1"	Performance	.67	.670
		Potential	.33	
	OER="3"	Performance	.60	.601
		Potential	.40	
All Subjects	OER="1"	Performance	.70	.702
		Potential	.30	
	OER="3"	Performance	.62	.702
		Potential	.38	

Mean individual  $R^2$  values were scattered about the .7 level, which is not considered unusually low or high, although other policy capturing models have resulted in somewhat higher  $R^2$  values (.75 and better);

Table 4.7  
Percentages of Individual Regressions Significant  
( $\alpha = .05$ )

PME Group	Total Number of Individual Regressions	OER Level	Number of Regressions Significant ( $\alpha = .05$ )	Percentage Significant
SOS	113	1	85	75.2%
	113	3	79	69.9%
ACSC	144	1	110	76.4%
	144	3	110	76.4%
AWC	120	1	85	70.8%
	120	3	86	71.6%

Glenn (1977) and O'Berry (1977) are examples of such models, although the policy capturing model involved was not concerned with employee appraisal in either case. Zedeck and Kafry report that previous studies of rater policies using policy capturing techniques yielded  $R^2$  values ranging from .50 to .80 (Zedeck and Kafry, 1977:275).

One hypothesis that was tested by this research was that officers use the same decision policies when making judgments of both "good" and "bad" OER ratings, and the data obtained from the individual regression models was used to test this hypothesis in a manner described in the next section of this chapter.

### Paired Sample t-Test of Individual Relative Weights

As outlined in earlier discussion, a paired sample t-test was conducted on the relative weights that officers assigned (implicitly) to predictors in the two scenario OER sets (e.g., a "1" or "3" OER rating). The null hypothesis tested in each case stated that the mean relative weight for each predictor was the same for both scenario OER levels. The alternate hypothesis stated that the mean relative weight for performance was greater in the OER=1 situation. Rejection of the null hypothesis therefore indicates that officers used different decision-making policies for judging "good" and "bad" OER ratings. The results of this analysis are presented in Table 4.8.

Table 4.8  
t-Test of Individual Relative Weights

Group	$t_0$	$t_{.05, \infty}$	Reject $H_0$ ?
SOS	2.33	1.645	Yes ( $p < .011$ )
ACSC	4.75	1.645	Yes ( $p < .001$ )
AWC	2.57	1.645	Yes ( $p < .006$ )
All Respondents	5.65	1.645	Yes ( $p < .001$ )

As noted in the table, the null hypothesis that the relative weights were the same, regardless of the scenario OER level, was rejected for all groups, with significance levels ranging from .011 to .001. This indicates that officers use different decision-making policies

concerning OER ratings which depend on the particular OER rating being judged.

### Group Regression Results

The analysis of group decision-making policies involved an approach similar to that used in the construction of individual regression models, the main difference being that the group regressions were based on the combined decisions made by the group members for each of the two OER ratings (e.g., 381 X 9 decisions for each OER level if all respondents are grouped together).

The SPSS subprogram "regression" was used in conducting group regressions: as in the regression of individual decisions, the appropriateness of assigned OER ratings was regressed on the two predictor variables "performance" and "potential."

Output from each regression included standardized regression coefficients (beta weights), the significance level of the model (all models were significant at the .001 level), ANOVA data used in later comparisons of regression models, and the  $R^2$  value for the group being modeled. As was pointed out in earlier discussion, relative weight is a concept that is meaningful only for individual decision-makers; thus, group regressions presented in this research do not have associated relative weights. Table 4.9 summarizes the results of the primary group regressions that were conducted. In order to provide a means of comparing group and individual regression models, both results are listed in the table.

Table 4.9  
Group Regression Results and a Comparison with  
Individual Regression Results

Group	Scenario OER (9 Scenarios for Each Score)	Predictor	Group Regressions		Individual Regressions	
			Standardized Beta Weight	Group $R^2$	Mean Relative Weight	Mean Ind. $R^2$
SOS Students	OER="1"	Performance	.647	.554	.70	.686
		Potential	.369		.30	
	OER="3"	Performance	-.601	.519	.64	.704
		Potential	-.398		.36	
ACSC Students	OER="1"	Performance	.687	.613	.73	.737
		Potential	.376		.27	
	OER="3"	Performance	-.592	.519	.62	.713
		Potential	-.413		.38	

Table 4.9 (continued)

Group	Scenario OER (9 Scenarios for Each Score)	Predictor	Group Regressions		Individual Regressions	
			Standardized Beta Weight	Group $R^2$	Mean Relative Weight	Mean Ind. $R^2$
AWC	OER="1"	Performance	.596	.486	.67	.670
		Potential	.361		.33	
	OER="3"	Performance	-.524	.432	.60	.601
		Potential	-.396		.40	
All Subjects	OER="1"	Performance	.645	.552	.70	.702
		Potential	.369		.30	
	OER="3"	Performance	-.570	.489	.62	.702
		Potential	-.401		.38	



Note that the beta weights for the OER="3" models are negative: this indicates, as one would expect, that the appropriateness of an OER rating of "3" was judged to be less as the levels of performance and potential increased. As in the individual regression models, performance was weighted more heavily than potential in the group models.

If one examines the tabled  $R^2$  values, it is noted that group  $R^2$ s are somewhat lower than mean individual  $R^2$ s. This results from the fact that although each decision maker may have made judgments with a fairly high degree of consistency, a model which is based on the total number of decisions made among all judges reflects a lower decision-making consistency ( $R^2$ ) due to differences in individual decision policies.

An interesting application of this  $R^2$  tendency in this research was that group  $R^2$ s for the OER="3" models were consistently lower than  $R^2$ s for the OER="1" models, although mean individual  $R^2$  values for the two OER levels were similar in magnitude. This suggests that although the consistency of individual judges was about the same for both OER levels, the lower group  $R^2$  values for the OER="3" model is indicative of more varied decision policies among officers making judgments of "3" OER ratings.

In order to determine whether the regression models for various groups of officers were different, F-test calculations were made, as described in the next section of this chapter.

### F-Test Comparison of Group Regression Models

The F-test calculation used to test whether group regression models being compared were significantly different is presented in Appendix D. In each comparison, the null hypothesis stated that there was no significant difference (at the .05 level) in the regression coefficients of the models being compared, versus the alternate hypothesis that there was at least one model whose regression coefficients were different. The results of the F-test comparisons are presented in Tables 4.10 through 4.13 based on groupings made by PME school, aeronautical rating, respondents' history as an OER rater, and respondents' most recent OER rating.

Table 4.10  
F-Test Comparison of Group Regression Models  
(PME School Groupings)

Groups Compared	Scenario OER	N*	F <sub>0</sub>	F <sub>.05</sub>	Reject H <sub>0</sub> ?
SOS/ACSC/AWC	"1"	3429	2.05	2.10	No
SOS/ACSC/AWC	"3"	3429	7.49	2.10	Yes
SOS/ACSC	"3"	2313	.68	2.60	No
SOS/AWC	"3"	2133	12.20	2.60	Yes
ACSC/AWC	"3"	2412	9.13	2.60	Yes
*N = number of decisions.					

Table 4.11  
F-Test Comparison of Group Regression Models  
(Aeronautical Rating)

Groups Compared	Scenario OER	N	$F_0$	$F_{.05}$	Reject $H_0$ ?
Pilot/Navigator/ Non-rated	"1"	3429	.853	2.10	No
Pilot/Navigator/ Non-rated	"3"	3429	1.525	2.10	No

Table 4.12  
F-Test Comparison of Group Regression Models  
(History as an OER Rater)

Groups Compared	Scenario OER	N	$F_0$	$F_{.05}$	Reject $H_0$ ?
Never/ Occasionally/ Frequently	"1"	3420	4.10	2.10	Yes
Never/ Occasionally/ Frequently	"3"	3420	5.376	2.10	Yes

As illustrated in Table 4.10, a comparison of regression models of the three PME schools yielded interesting results: the null hypothesis was not rejected when comparing the models with an assigned OER rating of "1," indicating that all groups used similar decision policies (i.e., used the performance and potential criteria in essentially the same manner). However, the null hypothesis was

Table 4.13  
F-Test Comparison of Group Regression Models\*  
(Recent OER Rating)

Groupings by Most Recent OER Rating	Scenario OER Level	$F_0$	$F_{.05}$	Reject $H_0$ ?
1/2/3	"1"	1.70	1.88	No
1/2/3	"3"	2.25	1.88	Yes
1/2&3	"3"	.162	2.6	No
1&2/3	"3"	3.86	2.6	Yes
*N = 3366 Decisions (374 Respondents)				

rejected when comparing the models where the assigned OER rating was a "3." This hypothesis rejection led to a further partitioning of the groups being compared: further analysis revealed that the AWC group had a decision-making model that was statistically different from the others.

There was no apparent difference in decision-making policies based on aeronautical rating, since the null hypothesis was not rejected.

When groupings were based on the recent (controlled) OER rating of respondents, the results were varied; there was no apparent difference in judgment policies among these officers when judgments of "1" ratings were made. However, when judgments of "3" ratings were made, the null hypothesis was rejected. Further analysis indicated

that the group of officers having received "3s" on their most recent (controlled) OER used significantly different decision policies when judging the appropriateness of "3" OER ratings in the decision-making exercise.

These results indicate that these officers' decision-making policies were, in some cases, significantly different on the basis of (1) history as an OER rater and (2) the respondent's recent OER rating, depending on the OER rating being judged. When the OER score being judged was a "3," the AWC group regression was significantly different than the models for SOS or ACSC. These differences do not necessarily suggest a causal relationship between these variables (e.g., a low OER rating causes a certain decision-making policy). However, by examining the standardized beta weights listed in Tables 4.14 and 4.15, one can observe the relative influence that these groups placed on the performance and potential factors.

Table 4.14  
Group Regression Coefficients\*  
(History as an OER Rater)

Group	Scenario OER="1"		Scenario OER="3"	
	$\beta_1$ (Performance)	$\beta_2$ (Potential)	$\beta_1$ (Performance)	$\beta_2$ (Potential)
Never Rated Another	.664	.348	-.612	-.378
Occasionally	.659	.376	-.560	-.422
Frequently	.619	.383	-.545	-.405
*Standardized				

Table 4.15  
Group Regression Coefficients\*  
(Recent OER Rating)

Group OER Rating	Scenario OER="1"		Scenario OER="3"	
	$\beta_1$ (Performance)	$\beta_2$ (Potential)	$\beta_1$ (Performance)	$\beta_2$ (Potential)
"1"	.652	.384	-.573	-.411
"2"	.639	.355	-.555	-.427
"3"	.651	.342	-.606	-.345
*Standardized				

Relative to groupings based on one's history as an OER rater, the analysis indicates that officers who have more frequently rated others weight potential most heavily. This is probably true because there is a high correlation between frequency as a rater and an officer's grade, coupled with the fact that the AWC group (senior officers) weighted potential most heavily.

When groupings were made by the recent (controlled) OER rating of respondents, the beta weights indicate that those officers having received the lowest OER ratings ("3s") weighted performance most heavily, for the scenario OER="3" model.

In summary, group regressions indicated that the AWC group weighted potential most heavily (compared to the other PME groups). Additionally, officers who have frequently been OER raters also weighted potential most heavily. However, officers who received

recent OER ratings of "3" weighted performance most heavily, and groupings based on aeronautical rating did not result in significantly different weightings among officers.

#### Examination of an Interactive Model

As explained in earlier discussion, this research was primarily concerned with judgment modeling based on an additive effect of predictor variables. However, an interaction term was added to the group regression models in order to examine the "strength" of the interaction between performance and potential predictors. The interaction term (the performance level times the potential level) was added to each regression model in the final regression step, and the significance and  $R^2$  increases were noted. The results of this analysis are presented in Table 4.16, as compared with the non-interactive model.

The significance level of the interaction term was .001 in all cases, although the  $R^2$  increases appeared to vary widely depending on the scenario OER level. For all groups, the interaction term added more explanative power to the regression models involving "good" OER ratings than it did to the models involving "bad" ratings. In essence, there appears to have been greater interaction between performance and potential criteria when decisions about "good" OERs were made. Since the  $R^2$  values for the OER="3" models did not increase appreciably when the interaction term was added, it appears

Table 4.16  
Comparison of Interactive and Non-interactive Models

Group	Scenario OER	Group R <sup>2</sup> (No Interaction)	Group R <sup>2</sup> (With Interaction)	% Increase in R <sup>2</sup>
SOS	1	.555	.616	11.0%
	3	.519	.526	1.3%
ACSC	1	.613	.673	9.8%
	3	.519	.529	.2%
AWC	1	.486	.553	13.8%
	3	.432	.438	1.4%
All Respond- ents	1	.552	.615	11.4%
	3	.486	.489	.6%

that decisions about "bad" OER ratings were based almost totally on the additive effect of the predictor variables.

### Part III Exercise Analysis

Part III of the exercise was included to obtain more direct reactions to OER scores with which the results of the policy capturing model could be compared. The mean values of "agreement" with the four commonly held beliefs about OER ratings are listed in Table 4.17 for each of the four OER situations (a "1" indicates total disagreement, and a "5" indicates total agreement).



Table 4.17  
Part III Responses  
(Mean Values and Standard Deviation)

"Belief"	Situation							
	Give a "1"		Receive a "1"		Give a "3"		Receive a "3"	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Absolute Performance	4.74	.51	4.69	.55	4.14	.94	4.00	1.08
Relative Performance	4.72	.63	4.59	.66	4.33	.84	4.20	.97
Promotion Potential	4.72	.58	4.67	.56	3.49	1.18	3.87	1.18
Effectiveness at Higher Levels	4.67	.54	4.49	.69	3.69	1.05	3.86	1.12

It is noted that the mean scores for "performance" beliefs were generally higher than the mean scores for "potential" beliefs. In order to test whether the performance scores were statistically different from the potential scores, a paired sample t-test was conducted on the mean values of aggregate performance and potential scores for each OER situation.

For each OER situation the agreement scores for the two performance beliefs were added and compared to a similarly constructed score for the two potential beliefs. The t-test tested the null hypothesis that the means of the two scores were the same, versus the alternate hypothesis that the performance score was higher. The results of these t-tests are presented in Table 4.18.

Table 4.18  
Paired Sample t-Test\*  
(Part III Data)

Situation	$t_0$	$t_{.05}$	Reject $H_0$ ?	Probability
Give a "1"	.96	1.645	No	$p < .169$
Receive a "1"	2.30	1.645	Yes	$p < .011$
Give a "3"	13.89	1.645	Yes	$p < .001$
Receive a "3"	6.19	1.645	Yes	$p < .001$
*N = 377				

In most cases, the null hypothesis was rejected; however, in the situation where an individual has (hypothetically) given a "1," the null hypothesis could not be rejected.

The general indication of this analysis is that this stronger agreement with performance beliefs is consistent with the findings of the policy capturing model, although there was strong agreement with the stated beliefs in all situations and therefore not much variance in the Part III responses.

## V Summary and Conclusions

The managerial activity known as employee appraisal has historically been viewed by many as a hindrance rather than an aid to organizational goal accomplishment. This negative view of appraisal systems may result from many factors, one of which is the subjectivity inherent in a system in which an individual's past and/or future worth to an organization is evaluated by another individual (or individuals).

The primary objective of employee appraisal is to improve results, i. e. facilitate organizational goal accomplishment. Many appraisal systems are structured such that the sole emphasis is on the appraisal of past accomplishments during a rating period, i. e. performance appraisal. However, some managerial appraisal systems are designed to evaluate individual potential in addition to one's performing during a reporting period.

The evaluation of potential involves an estimation of an individual's future worth to an organization, which subsequently forms an information base upon which job assignment, promotion, career progression, and other personnel administration decisions are based. Since potential evaluation is future-oriented, the criteria used in making such assessments of managers appear to result from individual perceptions of what factors should be considered. There is general

agreement that one's level of potential is indicated in part by past performance; however, potential assessment theoretically involves more than a de facto appraisal of performance. The appraisal of potential generally involves an examination of trends in past performance, coupled with an assessment of personality traits and individual capabilities; the frame of reference is generally with respect to some future type of job of expanded and/or different responsibilities.

Due to the interrelated nature of performance and potential, as well as the difficulties and varied perceptions involved in assessing these two concepts, there is an everpresent possibility that individuals' concepts of performance and potential could become confused and overlap significantly when judgments of appraisal ratings are made, i. e., potential is perceived to be heavily (if not totally) determined by one's performance during a rating period. If the assessment of potential is based on widely varied perceptions (among those involved with the appraisal system) of the criteria that should be used, then organizational dysfunctions may result. More specifically, these varied perceptions among raters and ratees may lead to conflict or frustrations which adversely affect the subordinate's performance, and hence organizational goal accomplishment.

The Air Force officer evaluation system is one appraisal system that requires an evaluation of potential, and the process is therefore subject to officers' varied perceptions of what criteria should be used to rate potential. The degree to which factors other than past

performance determine this rating of potential is not outlined in Air Force policy, hence the decision is made by each individual officer.

This research effort involved an examination of the relative influence that selected performance and potential criteria have on officers' judgments of OER ratings, particularly since some writers have suggested that the evaluation of potential may be little more than an evaluation of past performance. Since the judgment policies of Air Force officers are a result of individual perceptions, modeling these judgment policies was the primary method used in this research for determining the influence of performance and potential criteria on the judgment of OER ratings.

The methodology used in this research effort involved what is known as judgment modeling, or "policy capturing." Since the officer evaluation system involves judgments of officers' potential, the policy capturing technique was employed to construct a mathematical model of the process by which these judgments were made. The regression models constructed in this research were primarily based on the assumption that human judgment, although itself a very complex process, can be effectively modeled as a linear combination of stimulus variables, or cues. In this study, the decision variable was the "degree of appropriateness" of an OER rating assigned to a hypothetical officer, and the two predictor variables were indicators of performance and potential. Respondents to the decision-making exercise were asked to judge the appropriateness of OER ratings given to 18

hypothetical officers, each of whom was described in terms of different levels of performance and potential attributes. The exercises that were returned by 381 PME school students were used to construct both individual and group regression models which described the decision-making process of Air Force officers in making judgments of OER ratings.

### Conclusions

The various analysis results presented in chapter four of this report are related at this point to the stated research hypotheses. In each case, it is noted whether or not the null hypothesis was rejected, and the data analysis which supports the rejection (or non-rejection) is identified.

H1: Air Force officers combined performance and potential criteria in an additive rather than interactive manner when making judgments of OER ratings.

Group regression models which included an interaction term (performance level times the potential level) resulted in statistically significant interaction terms, although the percentage increase in  $R^2$ , i.e., the added explanative power of the model, was not considered substantial enough to question the value of the linear models in describing officers' judgment policies. The interaction term did lead to higher percentage increases in  $R^2$  values for the OER="1" models than the OER="3" models; however, the fact that non-interactive models accounted for 86.2% to 99.8% of the explained variance in the decision variable lends support to hypothesis H1.

H2: Air Force officers, when considering (1) job factors which are indicators of immediate past performance and (2) personal factors which are (other) indicators of potential, will make judgments of OER ratings (Section IV, "Evaluation of Potential") based equally on the job and personal factors.

For the performance and potential factors chosen, the group regression results presented in Table 4.8 show that the ratio of the beta weight for performance to the beta weight for potential ranged from a high of 1.83 to a low of 1.32. Additionally, the mean relative weights for the two predictors that were calculated from individual regression models indicated a fairly strong emphasis on performance: mean relative weights ranged from .73 (performance) and .27 (potential) to .60 (performance) and .40 (potential). These analysis results therefore lead to the rejection of hypothesis H2; i.e., officers place greater weight on the performance factor.

H3: Air Force officers use the same judgment policies when judging either "good" or "bad" OER ratings (i.e., "1" and "3" respectively) in terms of the performance and potential factors that influence these judgments.

This hypothesis was rejected, based on the results of the paired-sample t-test on the relative weights calculated in the individual regression models. This rejection indicates that officers use significantly different policies when judging good and bad OER ratings. Specifically, potential was weighted more heavily when officers made judgments of "3" ratings.

H4: Air Force officers are internally consistent in their judgment policies concerning OER scores.

The mean  $R^2$  values calculated in the individual regression models ranged from a low of .601 for the AWC students (scenario OER of "3") to a high of .737 for the ACSC students (scenario OER of "1"). This  $R^2$  value range is consistent with previous research regarding rater policies, and therefore supports hypothesis H3. It is noted, however, that the mean  $R^2$  values for the AWC students were the lowest of all the groups surveyed, particularly for the scenario OER equal to "3" models.

H5: Air Force officers represent a homogeneous group of decision makers regarding OER scores; i.e., use the same decision policies when making judgments of OER ratings, regardless of data groupings by PME school, aeronautical rating, OER rating history, and whether or not officers have been OER raters.

Based on the group regression models for the scenario OER of "1," this hypothesis was not rejected when comparisons were made on the basis of PME school or aeronautical rating, but the hypothesis was rejected when groupings were made on the basis of respondents' history as a rater.

For the group regression models with the scenario OER of "3," the results were different. The hypothesis was rejected for all groupings of the respondents with the exception of aeronautical rating. Further analysis of the groupings by PME school indicated that the AWC group had a statistically different regression model than the other two PME groups for the OER="3" model. Note that the rejection of this hypothesis indicates that there was variability (sometimes



substantial) among officers' judgment policies; in cases where variability was great, this may be an indication of some confusion (or at least lack of agreement) among officers regarding what constitutes a rating of "potential."

H6: The relative weights for performance and potential criteria determined by the policy capturing model are consistent with the intuitive responses given in Part III of the exercise; i.e., modeled judgment policies are consistent with officers' intuitive reactions.

This hypothesis is supported by the results of the paired sample t-test of the performance versus potential "agreement scores" presented in Table 4.18. It is noted, however that this particular analysis was diluted somewhat by the strong agreement shown by respondents with all reactions presented in the exercise (see Table 4.17).

#### Implications of the Study

As can be seen from the results of this research and the conclusions drawn, the judgment policies of Air Force officers regarding OER ratings are quite varied: for all groups, the models were different depending on whether the OER ratings assigned were "good" or "bad." Additionally, although the consistency with which officers made judgments of the ratings assigned was considered average for this type of research, the mean individual  $R^2$  values for AWC students (Lieutenant Colonels and Colonels) were the lowest of all groups surveyed. Furthermore, the low group  $R^2$  values (compared to the other PME groups) indicated that there was a wider variety in the

decision-making policies of AWC students; i.e., for this exercise the AWC group was not a very homogeneous group of decision makers. Further support for the differences in the AWC group was provided by the F-tests on the group regressions, which singled out AWC as having significantly different regression models than the other two schools (the AWC group weighted potential more heavily for the model where the scenario OER rating was a "3").

#### A Final Comment

The original statement of the research problem indicated that there may be confusion among officers regarding the concepts of performance and potential. In an attempt to address this problem, one area that was investigated was the relative influence that performance and potential (as represented by three indicators each) had on officers' judgments of OER ratings (which are, by definition, ratings of potential). It is clear that there is no correct answer to the question of what emphasis should be placed on performance versus other indicators of potential; since Air Force policy does not address the issue, each individual officer makes judgments of OER ratings based on his or her perceptions of what criteria should be used to evaluate potential. This situation therefore assures a great amount of subjectivity in an appraisal process that is considered by many to be the critical determinant of every officer's future in the Air Force.

A more definitive indicator of this confusion between performance and potential was the variability among officers' judgment

policies regarding OER ratings: the rejection for some groups of hypothesis H5, which related to group decision-making homogeneity, is an indication of some confusion among officers regarding what constitutes "potential."

### Bibliography

- AFP 36-26. Evaluator's Handbook. Washington: Department of the Air Force, September 1977.
- AFR 36-10. Officer Evaluations. Washington: Department of the Air Force, September 1975.
- AFR 36-10. Officer Evaluations. Washington: Department of the Air Force, January 1978.
- Air Force Human Resource Laboratory, USAF Officer Evaluation Systems Review and Research Recommendations. Brooks AFB, Texas, 23 February 1971.
- Anderson, Harry B. "The Rating Game: Formal Job Appraisals Grow More Prevalent But Get More Criticism," The Wall Street Journal, LVII: 1 (23 May 1978).
- Ansari, Shahid L. "An Integrated Approach to Control System Design," Accounting, Organizations and Society, 2: 101-102 (1977).
- Blakelock, Ralph A. An Analysis of the Impact of the New OER System on the Officer Corps Using a Lewin-Based Model. MS thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, September 1977.
- Campbell, John P., M.D. Dunnette, E. E. Lawler, and K. E. Weick, Jr. Managerial Behavior, Performance and Effectiveness. New York: McGraw-Hill Book Company, 1970.
- Chow, Gregory C. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, 28: 581-605 (July 1960).
- Chruden, Herbert J. and Arthur W. Sherman, Jr. Personnel Management (Third Edition). Cincinnati: South-Western Publishing Company, 1968.
- Cummings, L. I. and Donald P. Schwab. Performance in Organizations. Glenview, Illinois: Scott, Foresman, and Company, 1973.

- Dale, Ernest and Alice Smith. "Now Report Cards for Bosses," New York Times (March 31, 1957).
- Drucker, Peter F. The Practice of Management. New York: Harper and Row, 1954.
- The Effective Executive. New York: Harper and Row, 1967.
- Dunne, Edward J. Jr. A Reasoned Approach to Officer Evaluation, AFIT Technical Report 77-7, Wright-Patterson AFB, Ohio: Air Force Institute of Technology, August, 1977.
- Freund, John E. Mathematical Statistics. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.
- Glenn, Philip. A Judgment Analysis Approach to Examining a "Whole Person" Concept of Air Force Promotions. MS thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December, 1977.
- Gooch, Laurance L. Policy Capturing With Local Models: The Application of the AID Techniques in Modeling Judgment. PhD dissertation. Austin: University of Texas, 1972.
- Hoffman, P. J. "The Paramorphic Representation of Clinical Judgment," Psychological Bulletin, 47: 116-131 (1960).
- "Assessment of the Independent Contributions of Predictors," Psychological Bulletin, 59: 77-80 (1962).
- Huse, Edgar F. "Performance Appraisal--A New Look," Personnel Administration, 30: 3-5, 16-19 (March-April 1967).
- Johnson, Clarence A., John Meehan, and Robert E. Wilkinson. Officer Effectiveness Report Development--1971-1972. Lackland AFB, Texas: AFHRL, September 1976 (AFHRL-TR-76-61).
- Jones, Kirk M., Lawrence S. Mannis, Lester R. Martin, Jay L. Summers and Gerald R. Wagner. Judgment Modeling for Effective Policy and Decision Making, Volume 1. Austin: University of Texas, August, 1975 (ADAO 33186).
- Lavanagh, M. J., A. C. MacKinney, and L. Wolins. "Issues in Managerial Performance: Multitrait-Multimethod Analyses of Ratings," Psychological Bulletin, 75: 34-49 (1971).
- Kellogg, Marion S. What to Do About Performance Appraisal. New York: American Management Association, 1965.

- Keyserling, Steven. Junior Officer Perceptions of the OER System. MS thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December, 1976 (AD AO 38530).
- Koontz, Harold. Appraising Managers as Managers. New York: McGraw-Hill Book Company, 1971.
- Laurent, Harry. "Research on the Identification of Management Potential," Predicting Managerial Success (April 1968).
- Likert, Rensis. "Motivational Approach to Management Development," Harvard Business Review, 75: 75-80 (July-August 1959).
- Mann, Floyd C. and James K. Dent. "The Supervisor: Member of Two Organizational Families," Harvard Business Review, 32: 105-112 (November-December 1954).
- Madden, Joseph M. An Application to Job Evaluation of a Policy-Capturing Model for Analyzing Individual and Group Judgment. Lackland AFB, Texas: 6570th Personnel Research Laboratory, May 1963 (PRL-IDR-63-15).
- Mayfield, Harold. "In Defense of Performance Appraisal," Harvard Business Review, 38: 81-87 (March-April 1960).
- McGregor, Douglas M. "An Uneasy Look at Performance Appraisal," Harvard Business Review, 35: 89-94 (May-June 1957).
- Naylor, James C. and Robert J. Wherry, Jr. "The Use of Simulated Stimuli and the "JAN" Technique to Capture and Cluster the Policies of Raters," Educational and Psychological Measurement, 25: 969-985 (1965).
- Nie, H. H., C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent. Statistical Package for the Social Sciences. New York: McGraw-Hill Book Company, 1975.
- O'Berry, Carl G. An Experimental Test for Goal Congruence in an Air Force Major Command. MS thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, September 1977.
- Patz, Alan L. "Performance Appraisal: Useful But Still Resisted," Harvard Business Review, 53: 74-80 (May-June 1975).
- Pigors, Paul and Charles A. Meyers. Personnel Administration. New York: McGraw-Hill Book Company, 1977.

Rhoades, John W., Richard L. Sula, and Donald M. Wells. Impact of the New OER System--1976. Maxwell AFB, Alabama: Air Command and Staff College, April 1976 (AD BO 11901).

Smith, P. C. "Behaviors, Results and Organizational Effectiveness: The Problem of Criteria," in M. D. Dunnette, ed., Handbook of Industrial and Organizational Psychology. Chicago: Rand-McNally, 1976.

Stolz, Robert K. Appraisal Interviews: A Common Sense Approach. New York: McKinney and Company, 1960.

Terry, George R. Supervisory Management. Homewood, Illinois: Richard D. Irwin, Inc., 1974.

Ward, Joseph H. Jr. "Comment on the Paramorphic Representation of Clinical Judgment," Psychological Bulletin, 59: 74-76 (1962).

Whisler, Thomas L. and Shirley F. Harper, eds. Performance Appraisal: Research and Practice. New York: Holt, Rinehart and Winston, 1962.

Zedeck, Sheldon and Ditsa Kafry. "Capturing Rater Policies for Processing Evaluation Data," Organizational Behavior and Human Performance, 18: 269-294 (April 1977).

APPENDIX A

Decision-Making Exercise



A DECISION MAKING EXERCISE  
FOR  
AIR FORCE OFFICERS

THIS IS NOT A QUESTIONNAIRE. It is a decision making exercise that is designed to determine how Air Force officers make decisions regarding employee appraisal systems. The information derived from this exercise will be used in a master's thesis effort by one of your fellow officers at the Air Force Institute of Technology.

Please cooperate in carefully completing all three parts of this exercise so that we may gain additional insights into the decision making patterns of Air Force officers. Your participation in this exercise will be kept strictly confidential.

Thank you!

PLEASE DO NOT PUT YOUR NAME ON THIS EXERCISE

USAF SCN 78-108  
(Expires 23 Sep 78)

### PRIVACY STATEMENT

In accordance with paragraph 30, AFR 12-35, the following information is provided as required by the Privacy Act of 1974:

a. Authority

(1) 5 U.S.C. 301, Departmental Regulations: and/or

(2) 10 U.S.C. 80-12, Secretary of the Air Force, Powers and Duties, Delegation by.

b. Principal purposes. The survey is being conducted to collect information to be used in research aimed at illuminating and providing inputs to the solution of problems of interest to the Air Force and/or DOD.

c. Routine Uses. The survey data will be converted to information for use in research of management related problems. Results of the research, based on the data provided, will be included in written master's thesis and may also be included in published articles, reports, or texts. Distribution of the results of the research, based on the survey data, whether in written form or orally presented, will be unlimited.

d. Participation in this survey is entirely voluntary.

e. No adverse action of any kind may be taken against any individual who elects not to participate in any or all of this survey.

Part I : BACKGROUND INFORMATION

Please circle the letter corresponding to the most appropriate answer for each of the following questions:

1. What is your present grade?
  - a. Second Lieutenant
  - b. First Lieutenant
  - c. Captain
  - d. Major
  - e. Lieutenant Colonel
  - f. Colonel
2. What aeronautical rating do you hold?
  - a. Pilot
  - b. Navigator
  - c. Flight Surgeon
  - d. Non-rated
  - e. Other
3. To what command are you currently assigned?
  - a. Alaskan Air Command
  - b. U.S. Air Force Academy
  - c. Aerospace Defense Command
  - d. U.S. Air Forces in Europe
  - e. Air Force Accounting and Finance Center
  - f. Air Force Logistics Command
  - g. Air Force Systems Command
  - h. Air Reserve Personnel Center
  - i. Air Training Command (now includes Air University)
  - j. Headquarters Air Force Reserve
  - k. Headquarters USAF
  - l. Air Force Communications Service
  - m. Air Force Data Automation Agency
  - n. Headquarters Command
  - o. Military Airlift Command
  - p. Pacific Air Forces
  - q. Strategic Air Command
  - r. Tactical Air Command
  - s. USAF Security Service
  - t. Air Force Military Personnel Center
  - u. Air Force Inspection and Safety Center
  - v. Air Force Audit Agency
  - w. Air Force Office of Special Investigations
  - x. Other (specify \_\_\_\_\_)
4. What is your highest level of education now?
  - a. Some high school (did not graduate)
  - b. High school graduate (no college)
  - c. Some college (no degree)
  - d. College degree
  - e. Graduate work (no master's degree)
  - f. Master's degree
  - g. Postgraduate work beyond master's (no doctorate)
  - h. Doctorate degree

5. How much total active federal military service have you completed?

- |                            |                            |
|----------------------------|----------------------------|
| a. Less than 1 year        | l. Between 11 and 12 years |
| b. Between 1 and 2 years   | m. Between 12 and 13 years |
| c. Between 2 and 3 years   | n. Between 13 and 14 years |
| d. Between 3 and 4 years   | o. Between 14 and 15 years |
| e. Between 4 and 5 years   | p. Between 15 and 16 years |
| f. Between 5 and 6 years   | q. Between 16 and 17 years |
| g. Between 6 and 7 years   | r. Between 17 and 18 years |
| h. Between 7 and 8 years   | s. Between 18 and 19 years |
| i. Between 8 and 9 years   | t. Between 19 and 20 years |
| j. Between 9 and 10 years  | u. 20 years or more.       |
| k. Between 10 and 11 years |                            |

6. Which one of the following do you consider yourself?

- a. American Indian
- b. Asian Origin
- c. Black
- d. Spanish Speaking Origin
- e. White (other than Spanish Speaking Origin)
- f. Other (specify\_\_\_\_\_)

7. Which one of the following best describes your attitude toward making the Air Force a career?

- a. Definitely intend to make the Air Force a career
- b. Most likely will make the Air Force a career
- c. Undecided
- d. Most likely will not make the Air Force a career
- e. Definitely do not intend to make the Air Force a career

THE NEXT FOUR QUESTIONS ARE CONCERNED WITH YOUR ATTITUDES TOWARD YOUR CURRENT AIR FORCE JOB. IF YOU ARE CURRENTLY A STUDENT, PLEASE ANSWER THESE QUESTIONS IN THE CONTEXT OF YOUR LAST NON-STUDENT JOB.

8. Choose the statement which best describes how well you like your job.

- |                           |                               |
|---------------------------|-------------------------------|
| a. I hate it              | e. I like it                  |
| b. I dislike it           | f. I am enthusiastic about it |
| c. I don't like it        | g. I love it                  |
| d. I am indifferent to it |                               |

9. Which one of the following shows how much of the time you feel satisfied with your job?
- a. All of the time
  - b. Most of the time
  - c. A good deal of the time
  - d. About half of the time
  - e. Occasionally
  - f. Seldom
  - g. Never
10. Which one of the following shows how you think you compare with other people?
- a. No one likes their job better than I like mine
  - b. I like my job much better than most people like theirs
  - c. I like my job better than most people like theirs
  - d. I like my job about as well as most people like theirs
  - e. I dislike my job more than most people dislike theirs
  - f. I dislike my job much more than most people dislike theirs
  - g. No one dislikes their job more than I dislike mine
11. Which one of the following best describes how you feel about changing your job?
- a. I would quit this job at once if I could
  - b. I would take almost any other job in which I could earn as much as I am earning now
  - c. I would like to change both my job and my occupation
  - d. I would like to exchange my present job for another one
  - e. I am not eager to change my job, but I would do so if I could get a better job
  - f. I cannot think of any jobs for which I would exchange
  - g. I would not exchange my job for any other
12. Which statement best describes your involvement as an OER rater?
- a. I have never written an OER on another officer
  - b. I have occasionally written OER's on other officers
  - c. I have frequently written OER's on other officers (i.e., usually supervise at least one other officer)

13. Do you currently supervise and subsequently write OER's on at least one other officer? (If you are currently a student, please answer in the context of your most recent non-student assignment)

- a. Yes
- b. No

14. What rating did you receive on your most recent controlled OER (reviewer's rating) ?

- |        |                     |   |
|--------|---------------------|---|
| a. "1" | c. "3"              | e. Other (specify_____)                   |
| b. "2" | d. "4," "5," or "6" | f. I have never received a controlled OER |

15. What rating did you receive on your second mos. recent controlled OER (reviewer's rating) ?

- |        |                     |  |
|--------|---------------------|--|
| a. "1" | c. "3"              | e. Other (specify_____)                    |
| b. "2" | d. "4," "5," or "6" | f. I have received only one controlled OER |

PLEASE CONTINUE WITH PART II ON NEXT PAGE

## PART II :

EXERCISE INSTRUCTIONS

On the following pages of this exercise are a number of scenarios which present a rater's observations of a subordinate officer currently being evaluated through the OER system. These observations are listed in two categories: Job Factors and Personal Factors. In some cases, no observations will be given in one or both categories. In these cases, assume that there were no distinguishing observations, either good or bad, related to the category(s) that was(were) omitted. Also assume that the Job and Personal Factors noted in each scenario are observations by the rater and have not necessarily been included as written statements on the OER form itself.

Following the listing of rater observations, a statement is made which identifies the OER rating that was assigned to the officer being rated. For the purpose of this exercise, assume that there was agreement among the rater, additional rater, and reviewer regarding the rating assigned.

INTERPRET ALL RATINGS IN THE CONTEXT OF THE MOST RECENT OER SYSTEM:

- (1) A "1" rating is controlled, with a maximum of 22% of officers being rated receiving such a "top block" rating.
- (2) "2's" through "6's" are not controlled.

For the purpose of this exercise, assume that the following rating percentages apply to all scenarios:

OER Rating	Percentage of Officers Receiving This Rating in This Review Group
1	22%
2	58%
3	20%
Remaining Blocks (4 through 6)	0%

Therefore, a "1" rating indicates that at least 78% of the officers being rated were rated below the ratee who received a "1." Similarly, a "3" rating means that at least 80% of the officers being rated were rated above the ratee who received a "3."

After the scenario has been presented, you are then asked to evaluate each situation based only on the information provided in the scenario and your perception of whether or not the OER rating assigned is appropriate for the situation described. Assume that none of the officers in these scenarios deserve an OER rating lower than a "3."

ALWAYS ASSUME THAT THE RATEE'S GRADE IS THE SAME AS YOUR OWN!

EXAMPLE . SCENARIO

Each case is presented to you in the following format:

OFFICER # 0

A. JOB FACTORS:

1. (either three observations are listed, or the statement
2. "No distinguishing observations, either good or bad, are noted")
- 3.

B. PERSONAL FACTORS:

- 1.
2. (either three observations are listed, or the statement
3. "No distinguishing observations, either good or bad, are noted")

C. An OER rating of \_\_\_\_\_ was given (either a "1" or "3" will be entered)

DECISION # 0

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

PLEASE COMPLETE ALL CASES, AS ALL CASES ARE DIFFERENT



OFFICER #1

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 1

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

\*\*\*\*\*

OFFICER # 2

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 2

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

OFFICER # 3

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 3

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

\*\*\*\*\*

OFFICER # 4

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 4

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

OFFICER # 5

A. JOB FACTORS:

1. Accomplished all assignments.
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 5

How appropriate is the OER rating that was given?(circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

\*\*\*\*\*

OFFICER # 6

A. JOB FACTORS:

1. Accomplished all assignments.
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "3" was given.

DECISION # 6

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

OFFICER # 7

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "3" was given.

DECISION # 7

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5  
Definitely NOT      Probably NOT      Cannot      Probably      Definitely  
Appropriate      Appropriate      Determine      Appropriate      Appropriate

\*\*\*\*\*

OFFICER # 8

A. JOB FACTORS:

1. Accomplished all assignments.
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 8

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5  
Definitely NOT      Probably NOT      Cannot      Probably      Definitely  
Appropriate      Appropriate      Determine      Appropriate      Appropriate

OFFICER # 9

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "3" was given.

DECISION # 9

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

\*\*\*\*\*

OFFICER # 10

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 10

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

OFFICER # 11

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 11

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

\*\*\*\*\*

OFFICER # 12

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "1" was given.

DECISION # 12

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

OFFICER #13

A. JOB FACTORS:

1. Did not accomplish all assignments.
2. Not efficient in the use of time and other resources.
3. On-the-job communications needed improvement.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "1" was given.

DECISION # 13

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

\*\*\*\*\*

OFFICER # 14

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 14

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

OFFICER # 15

A. JOB FACTORS:

No distinguishing observations, either good or bad, are noted.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 15

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------

\*\*\*\*\*

OFFICER # 16

A. JOB FACTORS:

1. Accomplished all assignments
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

1. Not very dedicated or enthusiastic.
2. Of average intelligence.
3. Not very personable or sensitive to people and their needs.

C. An OER rating of "3" was given.

DECISION # 16

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely <u>NOT</u> Appropriate	Probably <u>NOT</u> Appropriate	Cannot Determine	Probably Appropriate	Definitely Appropriate
--------------------------------------	------------------------------------	---------------------	-------------------------	---------------------------



OFFICER # 17

A. JOB FACTORS:

1. Accomplished all assignments.
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

1. Very dedicated and enthusiastic.
2. Highly intelligent.
3. Very personable and sensitive to people and their needs.

C. An OER rating of "1" was given.

DECISION # 17

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

\*\*\*\*\*

OFFICER # 18

A. JOB FACTORS:

1. Accomplished all assignments.
2. Efficient in the use of time and other resources.
3. On-the-job communications very well done.

B. PERSONAL FACTORS:

No distinguishing observations, either good or bad, are noted.

C. An OER rating of "1" was given.

DECISION # 18

How appropriate is the OER rating that was given? (circle one number)

1.....2.....3.....4.....5

Definitely NOT  
Appropriate

Probably NOT  
Appropriate

Cannot  
Determine

Probably  
Appropriate

Definitely  
Appropriate

### PART III      EXERCISE INSTRUCTIONS

In this section, you are asked to give your reaction to four hypothetical situations wherein you are either receiving (as a ratee) or awarding (as a rater) various OER ratings. Assume that the interpretation of a "1" or a "3" is the same as in Part II of this exercise, i.e. 22% of rated officers receive "1s," 58% receive "2s," and 20% receive "3s." Also assume that there is agreement among the rater, additional rater, and the reviewer concerning the rating assigned.

- A. You have just received an OER rating of "3." For each of the possible reactions listed below, circle the number which best describes your feeling toward each reaction.

1. They (my boss, my chain of command, the "system") don't think I should be promoted.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reaction	Reaction		Reaction	Reaction

2. They are saying that most other people performed their job better than I did mine.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reaction	Reaction		Reaction	Reaction

3. They don't think I could be effective in higher level jobs.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reaction	Reaction		Reaction	Reaction

4. They are saying I didn't do my job very well.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reaction	Reaction		Reaction	Reaction

B. You have just given an OER rating of "1" to another officer. For each of the possible reasons listed below, circle the number which best describes your feeling toward each reason.

1. The officer was an outstanding performer during this period.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reasoning	Reasoning		Reasoning	Reasoning

2. I think this officer should be promoted.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reasoning	Reasoning		Reasoning	Reasoning

3. I think this officer could do a very effective job at higher levels.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reasoning	Reasoning		Reasoning	Reasoning

4. This officer performed his/her job better than most people.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reasoning	Reasoning		Reasoning	Reasoning

C. You have just received an OER rating of "1." For each of the possible reactions listed below, circle the number which best describes your feeling toward each reaction.

1. They (my boss, my chain of command, the "system") think I could be very effective at higher level jobs.

1 .....	2 .....	3 .....	4 .....	5 .....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

2. They are saying that I performed my job better than most people.

1 .....	2 .....	3 .....	4 .....	5 .....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

3. They are saying that I did my job very well.

1 .....	2 .....	3 .....	4 .....	5 .....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

4. They think I should be promoted.

1 .....	2 .....	3 .....	4 .....	5 .....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not My</u>	<u>Not My</u>		My	My
Reaction	Reaction		Reaction	Reaction

D. You have just given an OER rating of "3" to another officer. For each of the possible reasons listed below, circle the number which best describes your feeling toward each reason.

1. The officer did not perform well during this period.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reasoning	Reasoning		Reasoning	Reasoning

2. I think this officer should not be promoted.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reasoning	Reasoning		Reasoning	Reasoning

3. I think this officer could not do an effective job at higher levels.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reasoning	Reasoning		Reasoning	Reasoning

4. Most other people performed their jobs better than this officer.

1.....	2.....	3.....	4.....	5.....
Definitely	Probably	Unsure	Probably	Definitely
<u>Not</u> My	<u>Not</u> My		My	My
Reasoning	Reasoning		Reasoning	Reasoning

APPENDIX B

Air Force Form 707

I. RATEE IDENTIFICATION DATA <i>(Read AFR 36-10 carefully before filling in any item)</i>						
1. NAME <i>(Last, First, Middle Initial)</i>	3. SSN <i>(Include suffix)</i>		8. PERIOD OF REPORT			
2. ORGANIZATION, COMMAND, LOCATION, PAS CODE	4. PAFSC	5. DAFSC	FROM:			
			THRU:			
	6. ACTIVE DUTY GRADE		9. DAYS OF SUPERVISION			
	7. PERMANENT GRADE		10. REASON FOR REPORT			
II. JOB DESCRIPTION 1. DUTY TITLE: 2. KEY DUTIES, TASKS, AND RESPONSIBILITIES:						
III. PERFORMANCE FACTORS	NOT OBSERVED OR NOT RELEVANT	FAR 1 BELOW STANDARD	BELOW 1 STANDARD	MEETS STANDARD	ABOVE 1 STANDARD	WELL 1 ABOVE STANDARD
1. JOB KNOWLEDGE <i>(Depth, currency, breadth)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. JUDGMENT AND DECISIONS <i>(Consistent, accurate, effective)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. PLAN AND ORGANIZE WORK <i>(Timely, creative)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. MANAGEMENT OF RESOURCES <i>(Manpower and material)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. LEADERSHIP <i>(Initiative, accept responsibility)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. ADAPTABILITY TO STRESS <i>(Stable, flexible, dependable)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. ORAL COMMUNICATION <i>(Clear, concise, confident)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. WRITTEN COMMUNICATION <i>(Clear, concise, organized)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. PROFESSIONAL QUALITIES <i>(Attitude, dress, cooperation, bearing)</i>	0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. HUMAN RELATIONS <i>(Equal opportunity participation, sensitivity)</i>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



IV. RECOMMENDED ASSIGNMENT INFORMATION																											
1. STRONGEST QUALIFICATION:					3. ORGANIZATION LEVEL:																						
2. SUGGESTED JOB ASSIGNMENT (Include AFSC):					4. TIMING:																						
<div style="display: flex; justify-content: space-between;"> <div style="width: 60%;"> <p><b>V. EVALUATION OF POTENTIAL</b></p> <p style="text-align: center;"><b>TOP BLOCK CONTROLLED</b></p> <p><i>Evaluate the ratee's capability relative to that of officers in the same grade in the group being evaluated, for expanded/more diverse responsibility. Indicate your rating by placing an "X" in the designated portion of the appropriate block.</i></p> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="text-align: center;"> <table border="1" style="border-collapse: collapse;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> <p style="font-size: 8px;">RATER   ADDN RATER   REVR</p> </div> <div style="text-align: center;"> <table border="1" style="border-collapse: collapse;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> <p style="font-size: 8px;">RATER   ADDN RATER   REVR</p> </div> <div style="text-align: center;"> <table border="1" style="border-collapse: collapse;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> <p style="font-size: 8px;">RATER   ADDN RATER   REVR</p> </div> <div style="text-align: center;"> <table border="1" style="border-collapse: collapse;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> <p style="font-size: 8px;">RATER   ADDN RATER   REVR</p> </div> </div> </div> <div style="width: 35%; text-align: center;"> <table border="1" style="border-collapse: collapse; margin-bottom: 10px;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> <table border="1" style="border-collapse: collapse;"> <tr><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td><td style="width: 30px; height: 30px;"></td></tr></table> </div> <div style="position: relative; height: 100px; margin-top: 10px;"> <div style="position: absolute; right: -20px; top: 0; bottom: 0; text-align: center;"> Highest ↑ </div> </div> </div>																											

Lowest   ←

IF REPORT IS ABBREVIATED, "X" APPROPRIATE BLOCK

☐ REPORT CLOSING OUTSIDE CYCLE  
☐ RATEE SELECTED FOR PROMOTION

☐ RATEE RETIRES/SEPARATES WITHIN 4 MONTHS  
☐ LATE TO REVIEWER   ☐ OTHER (Rater explain)

APPENDIX C

Air Force Form 77

IDENTIFICATION DATA (Read AFM 36-10 carefully before filling out any item.)					
1. LAST NAME—FIRST NAME—MIDDLE INITIAL		2. AFSC		3. ACTIVE DUTY GRADE	
4. ORGANIZATION, COMMAND AND LOCATION		6. AERO RATING CODE		7. PERIOD OF REPORT FROM _____ THRU _____	
		8. PERIOD OF SUPERVISION		9. REASON FOR REPORT	
II. DUTIES—PAFSC _____ DAFSC _____					
III. RATING FACTORS (Consider how this officer is performing on his job.)					
1. KNOWLEDGE OF DUTIES					
NOT <input type="radio"/>	SERIOUS GAPS IN HIS KNOWLEDGE OF FUNDAMENTALS OF HIS JOB	SATISFACTORY KNOWLEDGE OF ROUTINE PHASES OF HIS JOB	WELL INFORMED ON MOST PHASES OF HIS JOB	EXCELLENT KNOWLEDGE OF ALL PHASES OF HIS JOB	EXCEPTIONAL UNDERSTANDING OF HIS JOB EXTREMELY WELL INFORMED ON ALL PHASES
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. PERFORMANCE OF DUTIES					
NOT <input type="radio"/>	QUALITY OR QUANTITY OF WORK OFTEN FAILS TO MEET JOB REQUIREMENTS	PERFORMANCE MEETS ONLY MINIMUM JOB REQUIREMENTS	QUANTITY AND QUALITY OF WORK ARE VERY SATISFACTORY	PRODUCES VERY HIGH QUANTITY AND QUALITY OF WORK MEETS ALL SUSPENSES	QUALITY AND QUANTITY OF WORK ARE CLEARLY SUPERIOR AND TIMELY
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. EFFECTIVENESS IN WORKING WITH OTHERS					
NOT <input type="radio"/>	INEFFECTIVE IN WORKING WITH OTHERS DOES NOT COOPERATE	SOMETIMES HAS DIFFICULTY IN GETTING ALONG WITH OTHERS	GETS ALONG WELL WITH PEOPLE UNDER NORMAL CIRCUMSTANCES	WORKS IN HARMONY WITH OTHERS A VERY GOOD TEAM WORKER	EXTREMELY SUCCESSFUL IN WORKING WITH OTHERS ACTIVELY PROMOTES HARMONY
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. LEADERSHIP CHARACTERISTICS					
NOT <input type="radio"/>	OFFEN WEAK FAILS TO SHOW INITIATIVE AND ACCEPT RESPONSIBILITY	INITIATIVE AND ACCEPTANCE OF RESPONSIBILITY ADEQUATE IN MOST SITUATIONS	SATISFACTORILY DEMONSTRATES INITIATIVE AND ACCEPTS RESPONSIBILITY	DEMONSTRATES A HIGH DEGREE OF INITIATIVE AND ACCEPTANCE OF RESPONSIBILITY	ALWAYS DEMONSTRATES OUTSTANDING INITIATIVE AND ACCEPTANCE OF RESPONSIBILITY
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. JUDGEMENT					
NOT <input type="radio"/>	DECISIONS AND RECOMMENDATIONS OFTEN WRONG OR INEFFECTIVE	JUDGEMENT IS USUALLY SOUND BUT MAKES OCCASIONAL ERRORS	SHOWS GOOD JUDGEMENT RESULTING FROM SOUND EVALUATION OF FACTORS	SOUND LOGICAL THINKER CONSIDERS ALL FACTORS TO REACH ACCURATE DECISIONS	CONSISTENTLY ARRIVES AT RIGHT DECISION EVEN ON HIGHLY COMPLEX MATTERS
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. ADAPTABILITY					
NOT <input type="radio"/>	UNABLE TO PERFORM ADEQUATELY IN OTHER THAN ROUTINE SITUATIONS	PERFORMANCE DECLINES UNDER STRESS OR IN OTHER THAN ROUTINE SITUATIONS	PERFORMS WELL UNDER STRESS OR IN UNUSUAL SITUATIONS	PERFORMANCE EXCELLENT EVEN UNDER PRESSURE OR IN DIFFICULT SITUATIONS	OUTSTANDING PERFORMANCE UNDER EXTREME STRESS MEETS THE CHALLENGE OF DIFFICULT SITUATIONS
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. USE OF RESOURCES					
NOT <input type="radio"/>	INEFFECTIVE IN CONSERVATION OF RESOURCES	USES RESOURCES IN A BARELY SATISFACTORY MANNER	CONSERVES BY USING ROUTINE PROCEDURES	EFFECTIVELY ACCOMPLISHES SAVINGS BY DEVELOPING IMPROVED PROCEDURES	EXCEPTIONALLY EFFECTIVE IN USING RESOURCES
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. WRITING ABILITY AND ORAL EXPRESSION					
NOT <input type="radio"/>	UNABLE TO EXPRESS THOUGHTS CLEARLY LACKS ORGANIZATION	EXPRESSES THOUGHTS SATISFACTORILY ON ROUTINE MATTERS	USUALLY ORGANIZES AND EXPRESSES THOUGHTS CLEARLY AND CONCISELY	CONSISTENTLY ABLE TO EXPRESS IDEAS CLEARLY	OUTSTANDING ABILITY TO COMMUNICATE IDEAS TO OTHERS
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IV. MILITARY QUALITIES (Consider how this officer meets Air Force standards.)					
NOT <input type="radio"/>	BEARING OR BEHAVIOR INTERFERES SERIOUSLY WITH HIS EFFECTIVENESS	CARELESS BEARING AND BEHAVIOR DETRACT FROM HIS EFFECTIVENESS	BEARING AND BEHAVIOR CREATE A GOOD IMPRESSION	ESPECIALLY GOOD BEHAVIOR AND BEARING CREATE A VERY FAVORABLE IMPRESSION	BEARING AND BEHAVIOR ARE OUTSTANDING HE EXEMPLIFIES THE MILITARY STANDARDS
OBSERVED <input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

AF FORM 77 PREVIOUS EDITION OF THIS FORM WILL BE USED UNTIL STOCK IS EXHAUSTED

COMPANY GRADE OFFICER EFFECTIVENESS REPORT

<b>V. OVER-ALL EVALUATION (Compare this officer ONLY with officers of the same grade.)</b>							
SPECIFIC JUSTIFICATION REQUIRED FOR THESE SECTIONS					SPECIFIC JUSTIFICATION REQUIRED FOR THESE SECTIONS		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
UNSATIS- FACTORY	MARGINAL	BELOW AVERAGE	EFFECTIVE AND COMPETENT		VERY FINE		EXCEPTIONALLY FINE
<b>VI. PROMOTION POTENTIAL</b>							
1 DOES NOT DEMONSTRATE A CAPABILITY FOR PROMOTION AT THIS TIME				<input type="checkbox"/>	2 PERFORMING WELL IN PRESENT GRADE SHOULD BE CONSID- ERED FOR PROMOTION ALONG WITH CONTEMPORARIES		<input type="checkbox"/>
3 DEMONSTRATES CAPABILITIES FOR INCREASED RESPONSIBILITY CONSIDER FOR ADVANCEMENT AHEAD OF CONTEMPORARIES				<input type="checkbox"/>	4 OUTSTANDING GROWTH POTENTIAL BASED ON DEMONSTRATED PERFORMANCE PROMOTE WELL AHEAD OF CONTEMPORARIES		<input type="checkbox"/>
<b>VII. COMMENTS</b>							
<b>VIII. REPORTING OFFICIAL</b>							
NAME, GRADE, AFSN, AND ORGANIZATION				DUTY TITLE		SIGNATURE	
				AERO RATING		CODE	
<b>IX. REVIEW BY INDORSING OFFICIAL</b>							
NAME, GRADE, AFSN, AND ORGANIZATION				DUTY TITLE		SIGNATURE	
				AERO RATING		CODE	

U. S. GOVERNMENT PRINTING OFFICE 1966 O7-238-329

APPENDIX D

F-Test Computational Form

## Appendix D

### Calculation of F-Test Values

The F-test values used to compare regression models in this study were calculated using the following formula:

$$F_0 = \frac{[SS_e - \sum_{j=1}^p SS_{ej}] / [(p-1)(K+1)]}{[\sum_{j=1}^p SS_{ej}] / [n-p(k+1)]}$$

where  $SS_e$  is the residual sum of squares derived by regressing all compared groups of decisions together,  $SS_{ej}$  is the residual sum of squares for the  $j$ th group of evaluations,  $p$  is the number of groups being compared (number of subsets of data in the regression),  $k$  is the number of predictor variables (two, in all cases for this study) and  $n$  is the total number of decisions in all groups being compared.

The null hypothesis being tested is

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p, \text{ where } \beta_i = \begin{matrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{matrix}$$

The alternate hypothesis is

$$H: \beta_i \neq \beta_j, \text{ for at least one } i, j \text{ pair.}$$

The null hypothesis is rejected if

$$F_0 > F_{\alpha}, [(p-1)(k+1)], [n-p(k+1)]$$

where  $\alpha = 0.05$  in all comparisons made for this study.

APPENDIX E

Demographic Frequencies



# EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 7/27/78)

Q1 GRADE

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

CODE

```

I
2. ***** ( 23)
I 1 LT
I
3. ***** ( 107)
I CAPT
I
4. ***** ( 127)
I MAJOR
I
5. ***** ( 95)
I LT COL
I
5. ***** ( 29)
I COLONEL
I
I.....I.....I.....I.....I.....I
1 40 80 120 160 200
FREQUENCY
    
```

MEAN	4.000	STD ERR	.53	MEDIAN	3.976
MODE	4.000	STD DEV	1.639	VARIANCE	1.079
KURTOSIS	-.649	SKEWNESS	.185	RANGE	4.000
MINIMUM	2.000	MAXIMUM	6.000	SUM	1524.000
C.V. PCT	25.958	.95 C.I.	3.895	TO	4.115
VALID CASES	381	MISSING CASES			

# EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 1/27/70)

Q2 AERO RATING

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

CONF

```

I
1. ***** ( 154)
I PILOT
I
2. ***** ( 49)
I NAVIGATOR
I
4. ***** ( 158)
I NON-RATED
I
I.....I.....I.....I.....I.....I
U      49      86      120      160      200
FREQUENCY
    
```

MEAN	2.451	STD DEV	.172	MEDIAN	2.441
MODE	4.000	STD DEV	1.413	VARIANCE	1.996
KURTOSIS	-1.885	SKENESS	.111	RANGE	3.000
MINIMUM	1.000	MAXIMUM	4.000	SUM	934.000
C.V. PCT	57.626	.95 C.I.	2.319	TO	2.594
VALID CASES	381	MISSING CASES	0		

## EXERCISE FREQUENCIES

FILE VONAME (OPERATION DATE = 07/27/78 )

Q3 COMMAND

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
AFA	2.	1	.3	.3	.3
ADCOM	3.	4	1.0	1.0	1.3
USAFE	4.	12	3.1	3.1	4.5
AFLO	6.	2	.5	.5	5.0
AFSC	7.	12	3.1	3.1	8.1
ATC	9.	25	6.6	6.6	73.8
HQ USAF	11.	4	1.0	1.0	74.8
AFCS	12.	3	.8	.8	75.6
MAC	15.	23	6.0	6.0	81.6
PACAF	16.	3	.8	.8	82.4
SAC	17.	30	7.9	7.9	90.3
TAC	18.	2	.5	.5	90.8
USAFSS	19.	2	.5	.5	91.3
AFMPC	20.	1	.3	.3	91.6
AFISC	21.	1	.3	.3	91.9
AFOSI	23.	1	.3	.3	92.2
OTHER	24.	4	1.0	1.0	93.2
TOTAL		381	100.0	100.0	

EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 7/27/78)

Q4 EDUCATION

CODE

```

I
2. * ( 1)
I HIGH SCHOOL
I
3. ** ( 2)
I SOME COLLEGE
I
4. ***** ( 51)
I COLLEGE DIPLOMA
I
5. ***** ( 64)
I GRADUATE WORK
I
6. ***** ( 195)
I MASTERS
I
7. ***** ( 37)
I POSTGRADUATE
I
8. ***** ( 21)
I DOCTORATE
I
I.....I.....I.....I.....I.....I.....I.....I
0          41          80          120          160          200
FREQUENCY

```

MEAN	5.693	STD ERR	.154	MEDIAN	5.821
MODE	5.911	STD DEV	1.161	VARIANCE	1.124
KURTOSIS	.176	SKENESS	-.134	RANGE	6.000
MINIMUM	2.000	MAXIMUM	8.000	SUM	2169.000
C.V. PCT	18.622	.95 C.I.	5.186	TD	5.800

VALID CASES 381 MISSING CASES 0

EXERCISE FREQUENCIES

FILE NONAME (ORIENTATION DATE = 7/27/78)

Q5 YEARS SERVICE

CATEGORY LABEL	CODE	ABSOLUTE FREQ	RELATIVE FREQ (PCT)	ADJUSTED FREQ (PCT)	CUM FREQ (PCT)
2 TO 3	3.	3	.8	.8	.8
3 TO 4	4.	11	2.9	2.9	3.7
4 TO 5	5.	17	4.5	4.5	8.2
5 TO 6	6.	21	5.5	5.5	13.7
6 TO 7	7.	19	5.0	5.1	18.7
7 TO 8	8.	21	5.5	5.5	24.2
8 TO 9	9.	10	2.6	2.6	26.8
9 TO 10	10.	3	.8	.8	27.6
10 TO 11	11.	12	3.1	3.2	30.8
11 TO 12	12.	42	11.0	11.1	41.8
12 TO 13	13.	39	10.2	10.3	52.1
13 TO 14	14.	31	8.1	8.2	60.3
14 TO 15	15.	12	3.1	3.2	63.4
15 TO 16	16.	8	2.1	2.1	65.5
16 TO 17	17.	8	2.1	2.1	67.6
17 TO 18	18.	29	7.6	7.6	75.3
18 TO 19	19.	29	7.6	7.6	82.9
19 TO 20	20.	27	7.1	7.1	90.0
20 OR MORE	21.	38	10.0	10.0	100.0
		1	.3	MISSING	
TOTAL		381	100.0	100.0	

EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 11/27/78)

Q6 RACE

```

CODE
1. * ( 1)
   I AMER INDIAN
   I
2. * ( 4)
   I ASIAN
   I
3. ** ( 9)
   I BLACK
   I
4. * ( 2)
   I SPANISH
   I
5. ***** ( 359)
   I WHITE
   I
6. ** ( 5)
   I OTHER
   I
(MISSING) * ( 1)
          I
          I
          I.....I.....I.....I.....I.....I
          0          1          2          3          4          5
          FREQUENCY

```

MEAN	4.918	STD ERR	.125	MEDIAN	4.985
MODE	5.000	STD DEV	.494	VARIANCE	.244
KURTOSIS	26.521	SKENNESS	-4.412	RANGE	5.000
MINIMUM	1.000	MAXIMUM	6.000	SUM	1869.000
C.V. PCT	10.343	.95 C.I.	4.659	TD	4.958
VALID CASES	380	MISSING CASES	1		

EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 07/27/78 )

07 CAREER INTENT

000F

```

I
1. ***** ( 257)
I DEFINITELY
I
2. ***** ( 75)
I MOST LIKELY
I
3. *** ( 23)
I UNDECIDED
I
4. *** ( 13)
I PROBABLY NOT
I
5. ** ( 9)
I DEFINITELY NOT
I
* ( 2)
(MISSING) I
I
I.....I.....I.....I.....I.....I.....I
0 1 2 3 4 5
FREQUENCY

```

MEAN	1.533	STD ERP	.49	MEDIAN	1.237
MODE	1.000	STD DEV	.646	VARIANCE	.895
KURTOSIS	3.542	SKEWNESS	2.115	RANGE	4.000
MINIMUM	1.000	MAXIMUM	5.000	SUM	581.716
C.V. PCT	51.715	.95 C.I.	1.437	TD	1.629
VALID CASES	379	MISSING CASES	2		

EXERCISE FREQUENCIES

FILE NO NAME (CREATION DATE = 11/27/78)

08 HOPPOCK 1

CODE

```

1. ** ( 4)
I
I
2. *** ( 7)
I
I
3. ***** ( 23)
I
I
4. **** ( 12)
I
I
5. *****( 174)
I
I
6. *****( 145)
I
I
7. *****( 54)
I
I
* ( 1)

```

(MISSING)

```

I.....I.....I.....I.....I.....I.....I
4 8 12 16 20
FREQUENCY

```

MEAN	5.418	STD ERR	.060	MEDIAN	5.568
MODE	6.000	STD DEV	1.179	VARIANCE	1.359
KURTOSIS	2.91	SKEWNESS	-1.210	RANGE	6.000
MINIMUM	1.000	MAXIMUM	7.000	SUM	259.000
C.V. PCT	21.752	.95 C.I.	5.200	TO	5.537
VALID CASES	380	MISSING CASES	1		



EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 17/27/78)

Q9 HOPPOCK 2

CODE

```

1. *** ( 17)
I
I
2. ***** ( 227)
I
I
3. ***** ( 71)
I
I
4. **** ( 30)
I
I
5. **** ( 33)
I
I
6. * ( 5)
I
I
7. * ( 1)
I
I
I.....I.....I.....I.....I.....I
  1 2 3 4 5
FREQUENCY

```

MEAN	2.648	STD ERR	.57	MEDIAN	2.28
MODE	2.000	STD DEV	1.119	VARIANCE	1.231
KURTOSIS	1.219	SKENNESS	1.351	RANGE	6.000
MINIMUM	1.000	MAXIMUM	7.000	SUM	1016.000
C.V. PCT	42.318	.95 C.I.	2.129	TO	2.752
VALID CASES	381	MISSING CASES	0		

EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 07/27/76)

Q10 HOPPOCK 3

0000

```

1. *** ( 5)
I
I
2. *****( 105)
I
I
3. *****( 154)
I
I
4. *****( 97)
I
I
5. *****( 13)
I
I
6. ** ( 4)
I
I
(MISSING) ( 1)
I
I
I.....I.....I.....I.....I.....I.....I

```

FREQUENCY 4 8 12 16 20

MEAN	3.053	STD. ERR.	.148	MEDIAN	3.013
MODE	3.000	STD. DEV.	.931	VARIANCE	.867
KURTOSIS	.154	SKENNESS	.426	RANGE	8.000
MINIMUM	1.000	MAXIMUM	6.000	SUM	1151.000
C.V. PCT	30.393	.95 C.I.	2.059	TD	3.157
VALID CASES	380	MISSING CASES	1		

EXERCISE FREQUENCIES

FILE NONAME (CREATION DATE = 07/27/78)

Q11 HOPPOCK L

```

0005
1. *** ( 3)
   I
   I
2. ** ( 5)
   I
   I
3. ***** ( 18)
   I
   I
4. *****( 14)
   I
   I
5. *****( 193)
   I
   I
6. *****( 48)
   I
   I
7. ** ( 5)
   I
   I
(MISSING) ** ( 2)
          I
          I
          I.....I.....I.....I.....I.....I.....I
          0          40          80          120          160          200
          FREQUENCY

```

MEAN	4.641	STD ERR	.152	MEDIAN	4.777
MODE	5.000	STD DEV	1.004	VARIANCE	1.008
KURTOSIS	3.215	SKELNESS	-1.172	RANGE	9.000
MINIMUM	1.000	MAXIMUM	7.000	SUM	1759.300
C.V. PCT	21.637	.95 C.I.	4.14	TD	4.743
VALID CASES	379	MISSING CASES	2		

# EXERCISE FREQUENCIES

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

FILE NONAME (OPERATION DATE = 07/27/78)

Q12 HISTORY AS A PATER

```

CODE
1. I ***** ( 117)
   I NEVER
   I
2. I ***** ( 112)
   I OCCASIONALLY
   I
3. I ***** ( 150)
   I FREQUENTLY
   I
J ** ( 2)
(MISSING) I
          I
          I.....I.....I.....I.....I.....I
          0      40      80      120      160      200
          FREQUENCY
    
```

MEAN	2.137	STD ERR	.043	MEDIAN	2.147
MODE	3.000	STD DEV	.636	VARIANCE	.679
KURTOSIS	-1.551	SKEWNESS	-.105	RANGE	2.
MINIMUM	1.000	MAXIMUM	3.000	SUM	731.000
C.V. PCT	40.52	.95 C.I.	2.113	TD	2.171
VALID CASES	779	MISSING CASES	2		

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

Q13 CURRENTLY SUPERVISE OFFICER?

```
0005  
I  
1. ***** ( 159)  
I YES  
I  
2. ***** ( 221)  
I NO  
I  
I * ( 1)  
(MISSING) I  
I  
I.....I.....I.....I.....I.....  
3          100          200          300          400          500  
FREQUENCY
```

MEAN	1.532	STD. DEV.	.25	MEDIAN	1.616
MODE	2.000	STD. DEV.	.694	VARIANCE	.244
KURTOSIS	-1.910	SKEWNESS	-.332	RANGE	1.000
MINIMUM	1.000	MAXIMUM	2.000	SUM	6.160
C.V. PCT	31.231	.95 C.I.	1.532	TO	1.631
VALID CASES	340	MISSING CASES	1		

EXERCISE FREQUENCIES

FILE NONAME (OPERATION DATE = 07/27/78)

014 RECENT DEP RATING

```

0000
1. ***** ( 192)
I 1
I
2. ***** ( 112)
I 2
I
3. ***** ( 76)
I 3
I
4. * ( 1)
I " , " , OP R
I
5. ** ( 2)
I OTHER
I
6. ** ( 3)
I NO DEP
I
" * ( 1)
(MISSING) I
I
I.....I.....I.....I.....I.....I.....I
0 4 8 12 16 20
FREQUENCY

```

MEAN	1.732	STD DEP	.66	MEDIAN	1.49
MODE	1.000	STD DEV	.897	VARIANCE	.814
KURTOSIS	3.148	SKENNESS	1.416	RANGE	5.000
MINIMUM	1.000	MAXIMUM	5.000	SUM	658.000
C.V. PCT	51.770	.95 C.I.	1.000	TO	1.822
VALID CASES	380	MISSING CASES	1		

# EXERCISE FREQUENCIES

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

FILE NONAME (CREATION DATE = 7/27/78)

Q15 SECOND MOST FRECENT CEE EATING

```

CODE
1. ***** ( 187)
   I 1
   I
2. ***** ( 11)
   I 2
   I
3. ***** ( 43)
   I 3
   I
5. ** ( 2)
   I OTHER
   I
6. ***** ( 38)
   I NO DEP OR ONLY ONE
   I
* ( 1)
(MISSING) I
          I
          I.....I.....I.....I.....I.....I.....I
          0         40         80         120         160         200
          FREQUENCY
    
```

MEAN	2.037	STD ERR	.177	MEDIAN	1.527
MODE	1.000	STD DEV	1.512	VARIANCE	2.257
KURTOSIS	2.123	SKENNESS	1.758	RANGE	5.000
MINIMUM	1.000	MAXIMUM	6.000	SUM	771.000
C.V. PCT	73.751	.95 C.I.	1.885	TO	2.158
VALID CASES	380	MISSING CASES	1		

### Vita

Drew W. Browning was born in Baton Rouge, Louisiana on October 12, 1949. After graduating from Baker High School in 1967, he attended Louisiana State University, majoring in chemical engineering. In May 1972, Captain Browning graduated from LSU with a bachelor of science degree in chemical engineering and received his commission in the United States Air Force through the Reserve Officer Training Corps program.

Captain Browning entered active duty in October 1972, and attended the 38 week Electronics Systems Officer Course at Keesler AFB, Mississippi. Upon graduation from technical school, he completed tours of duty as the communications-electronics-meteorological maintenance officer at the 645th Radar Squadron, Patrick AFB, Florida, and the 692nd Radar Squadron, Baudette AFS, Minnesota.

In May, 1977, Captain Browning reported to the Air Force Institute of Technology and was enrolled in the Graduate Systems Management program. Following graduation from AFIT, Captain Browning will be assigned to the Defense Warning Program Branch of Headquarters Air Force Communications Service, Scott AFB, Illinois.

#### PERMANENT ADDRESS:

11835 Cooper Drive  
Baton Rouge, Louisiana 70811



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GSM/SM/78S-3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE PERFORMANCE-POTENTIAL DILEMMA-- DOES IT EXIST?		5. TYPE OF REPORT & PERIOD COVERED M.S. Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Drew W. Browning Capt USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE September 1978
		13. NUMBER OF PAGES 151
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Approved for public release; IAW AFR 190-17  JOSEPH P. HIPPS, Major, USAF Director of Information  OCT 10 1978		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Potential Policy capturing Performance Rating criteria Appraisals OER		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  This research was conducted to examine Air Force officers' perceptions both individually and in groups, of what is involved in a rating of potential. In order to determine these perceptions, a judgment modeling, or policy capturing research approach was used. A decision-making exercise was administered to 381 active duty USAF officers attending Squadron Officers' School, Air Command and Staff College, and Air War College. In this exercise, officers		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

were asked to judge the appropriateness of OER ratings assigned to 18 hypothetical officers, based on different levels of two factors: (1) the officer's past performance, as illustrated by three indicators and (2) the officer's potential, as illustrated by three indicators other than past performance. Data collected from the exercise were then analyzed to test various hypotheses concerning the relative influence of performance and (other) potential factors on officers' judgments of OER ratings.

The results of the research indicate that Air Force officers do not view the OER rating as being solely determined by past performance, although the performance factor was weighted most heavily. Analysis indicated that officers used different judgment policies when considering "good" and "bad" OER ratings, and that they were generally internally consistent in their judgment policies. Additionally, several different analyses indicated that the Air War College students used judgment policies that were significantly different than the other groups; specifically, AWC students generally placed greater emphasis on potential than the other PME groups. Furthermore, the AWC group displayed both the lowest internal judgment consistency and the lowest group  $R^2$  values (an indication of the degree of decision-making homogeneity within the group).

←

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)